Journal of Hydrology: Regional Studies 30 (2020) 100710

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

Establishing uncertainty ranges of hydrologic indices across climate and physiographic regions of the Congo River Basin

Pierre M. Kabuya[a,b,*], Denis A. Hughes[a], Raphael M. Tshimanga[b], Mark A. Trigg[c], Paul Bates[d]

[a] Institute for Water Research, Rhodes University, Box 94, Grahamstown, 6140, South Africa
[b] Congo River Basin Water Resources Research Center (CRREBaC) and Department of Natural Resources Management, University of Kinshasa, Kinshasa, Congo
[c] School of Civil Engineering, University of Leeds, Leeds, LS2 9JT, UK
[d] School of Geographical Sciences, University of Bristol, Bristol, BS8 1SS, UK

ARTICLE INFO

Keywords:
Sub-basin classification
Congo River Basin
Hydrologic indices
Uncertainty ranges

ABSTRACT

Study region: The five drainage systems of the Congo River Basin in central Africa.
Study focus: This study aims to establish uncertainty ranges of hydrologic indices and to provide a basis for transferring hydrologic indices from gauged to ungauged sub-basins by identifying the most influential climate and physiographic attributes.
New insights for this region: Only limited information on individual sub-basins natural hydrology exists across the Congo River Basin, limiting the application of commonly used regionalization approaches for prediction in ungauged sub-basins. This study uses predictive equations for the hydrologic indices across all climate and physiographic regions based only on the aridity index. The degree of uncertainty in the derived uncertainty bounds is less than 41% for both Q10/MMQ and Q50/MMQ indices across the basin. A greater degree of uncertainty is associated with the runoff ratio and the Q90/MMQ indices. The uncertainty is assumed to be due to uncertainty in rainfall and evapotranspiration estimates, a lack of spatial representativeness of the available observed streamflow data and other factors (e.g., geology) that might control the hydrologic indices rather than the aridity index alone. The uncertainty ranges provide the first estimates of hydrologic indices that are intended to constrain the outputs from hydrologic models and appropriately quantify prediction uncertainty and risks associated with water resources decision making.

1. Introduction

Hydrologic indices or signatures are the characteristics of a sub-basin's long-term hydrologic behavior. They reflect the dynamics of the different components of the catchment water balance such as climate, water storage and different runoff processes. They have been used in many hydrologic applications such as directly for runoff prediction (Kult et al., 2014; Zhang et al., 2018), model evaluation and optimization (Shafii and Tolson, 2015), model selection (Jothityangkoon et al., 2001; McMillan et al., 2011), uncertainty analysis (Westerberg and Mcmillan, 2015; Westerberg et al., 2016), environmental flow assessment (Olden and Poff, 2003), catchment classification (Ley et al., 2011; Sawicz et al., 2011) and ensemble predictions (Yadav et al., 2007; Zhang et al., 2008;

* Corresponding author at: Institute for Water Research, Rhodes University, Box 94, Grahamstown, 6140, South Africa.
E-mail address: kabuyabal@gmail.com (P.M. Kabuya).

https://doi.org/10.1016/j.ejrh.2020.100710
Received 9 September 2019; Received in revised form 5 March 2020; Accepted 26 June 2020

Tumbo and Hughes, 2015; Hughes, 2016; Ndzabandzaba and Hughes, 2017). There exists a large range of possible indices that can be derived from different sources and used for a wide range of applications.

Olden and Poff (2003) provide a large range of possible indices. These include the flow distribution, event frequency and duration, flow dynamics, rainfall-runoff ratio and rainfall, and other climate-based indices. Flow distribution indices refer to the long-term mean flow and flow percentiles. The event frequency and duration indices cover high and low flow event frequencies and durations (Clausen and Biggs, 2000; Olden and Poff, 2003). The flow dynamics include the slope of the normalized flow duration curve (Yadav et al., 2007), overall flow variability, the base flow (Clausen and Biggs, 2000), low and high-flow variability and flow autocorrelation (Euser et al., 2013). Westerberg et al. (2016) used 15 hydrologic signatures including 9 for flow distribution (flow percentiles) and 6 for flow dynamics (base flow index, coefficient of variability in streamflow, etc…) for hydrologic similarity. Zhang et al. (2018) used 13 runoff signatures including 3 for low flows, 1 for high flows, 4 for mean annual and mean seasonal flows and 5 for flow dynamics for assessing prediction accuracy in ungauged sub-basins.

The use of hydrologic indices for constraining rainfall-runoff model outputs in ungauged sub-basins has been suggested as an alternative to traditional model calibration processes (Addor et al., 2018). In a pilot study using 30 watersheds in the United Kingdom, Yadav et al. (2007) used streamflow indices such as high pulse count, runoff ratio and the slope of the flow duration curve to constrain uncertainty ensemble outputs from a hydrologic model. Similarly, Zhang et al. (2008) used a multi-objective framework for identifying behavioral parameter ensembles for ungauged basins using suites of regionalized hydrologic indices and they concluded that regionalization of these streamflow characteristics provided an additional way to extrapolate information from gauged to ungauged sub-basins for use in continuous basin scale modeling. Shafii and Tolson (2015) used a large number of sub-basin response indices in a multi-objective optimization context to achieve a level of acceptability for each index in ungauged sub-basins. Thus, there has been a growing interest in the use of hydrologic indices in the context of ensemble model predictions to constrain the uncertainty of predictions in ungauged sub-basins.

This study builds on a growing recognition of the utility of hydrologic indices for constraining rainfall-runoff models in the data scarce regions of southern Africa (Tumbo and Hughes, 2015; Hughes, 2016; Ndzabandzaba and Hughes, 2017). A study in Tanzania by Tumbo and Hughes (2015**)** quantified six hydrologic constraints from observed streamflow data and regionalized them using a limited amount of sub-basin physical property (e.g. land cover, soil texture and topography) and climate data. Although the results were encouraging, there was a need to refine the constraint bounds, the input parameter sets and assess suitable climate input data in order to achieve a more robust model of the basin. Ndzabandzaba and Hughes (2017) used pre-existing model simulations for the quantification of indices and regionalized them based on the sub-basins' hydrologic similarities. The indices were then used to constrain the uncertainty ensemble outputs from a new model and the results were assessed against the available observed data (Ndzabandzaba and Hughes, 2017). All these studies used similar hydrologic constraints, namely the mean monthly runoff volume (MMQ in $m^3 * 10^6$), mean monthly groundwater recharge depth (MMR in mm), the 10th, 50th and 90th percentiles of the flow duration curve expressed as a fraction of MMQ (Q10/MMQ, Q50/MMQ, Q90/MMQ) and the percentage of time that zero flows are expected (%Zero). These were judged to be the minimum number of key indices that can discriminate between different hydrologic responses (Ndzabandzaba and Hughes, 2017).

In the Congo River Basin, despite its major contribution to regional and global ocean circulation and climate (Santini and Caporaso, 2018; Spencer et al., 2016; Dargie et al., 2017; Dyer et al., 2017), less is known about its hydrology (Alsdorf et al., 2016) compared to other large tropical basins such as Amazon and Orinoco (Wongchuig et al., 2017). Some recent research has investigated the climate of the basin (Bell et al., 2015; Dyer et al., 2017; Ndehedehe et al., 2019), the biogeochemistry (Spencer et al., 2009, 2010; Spencer et al., 2016), changes in the forest cover (Mayaux et al., 2000; Hansen et al., 2008; Somorin et al., 2012), the impacts of wetlands (Bwangoy et al., 2010; Lee et al., 2011, 2015; Yuan et al., 2017; Becker et al., 2018), basin-scale hydrologic modeling (Beighley et al., 2011; Tshimanga and Hughes, 2012, 2014; Aloysius and Saiers, 2017), hydrodynamic modeling (Kabuya et al., 2018; O'Loughlin et al., 2019), soil erosion and sediment production (Coynel et al., 2005; Kabantu et al., 2018; Mushi et al., 2019) and river bathymetry and water level changes (Bos et al., 2006; O'Loughlin et al., 2013; Kim et al., 2017; Carr et al., 2019). Alsdorf et al. (2016), reviewed the scope of some of these studies and identified the further research opportunities that exist.

Tshimanga and Hughes (2014) identified the main sources of uncertainty in the application of hydrologic models in the Congo River Basin and recommended the use of regionalized hydrologic indices to overcome some of the problems of data scarcity and to reduce uncertainty in hydrologic simulations. Their application to ungauged sub-basins requires a thorough understanding of their likely ranges, which are likely to be highly variable across climate and physiographic regions. Indices, such as long-term mean monthly flow volume and percentiles of the flow duration curve, can be readily obtained from observed streamflow data. However, there are relatively few stations within the data scarce Congo River Basin and many of those that do exist are located on large rivers that represent heterogeneous upstream responses, and therefore are not useful to quantify regional patterns of response. In addition, many of the available streamflow time series are short and the data from different stations rarely coincide in time, representing different sequences of dry and wet flow conditions such that the derived indices at different stations may not be comparable with each other.

The key objective of this study is to establish uncertainty ranges of hydrologic indices across the Congo River Basin. Specifically, the study focuses on (i) providing a basis for extrapolating the hydrologic indices from gauged to ungauged sub-basins (ii) identifying potential predictors of hydrologic behavior and (iii) quantifying uncertainty ranges of the hydrologic indices. Due to the largely ungauged nature of the Congo River Basin, catchment classification offers an approach for reducing the complexity of the basin to a few groups of sub-basins where the differences in climate and physiographic characteristics (and hence the hydrologic indices) are assumed to be greater between the groups than within each group. The uncertainty ranges of the indices for each group are designed to reflect not only the internal homogeneity within the group, but also our lack of knowledge associated with the limited amount of

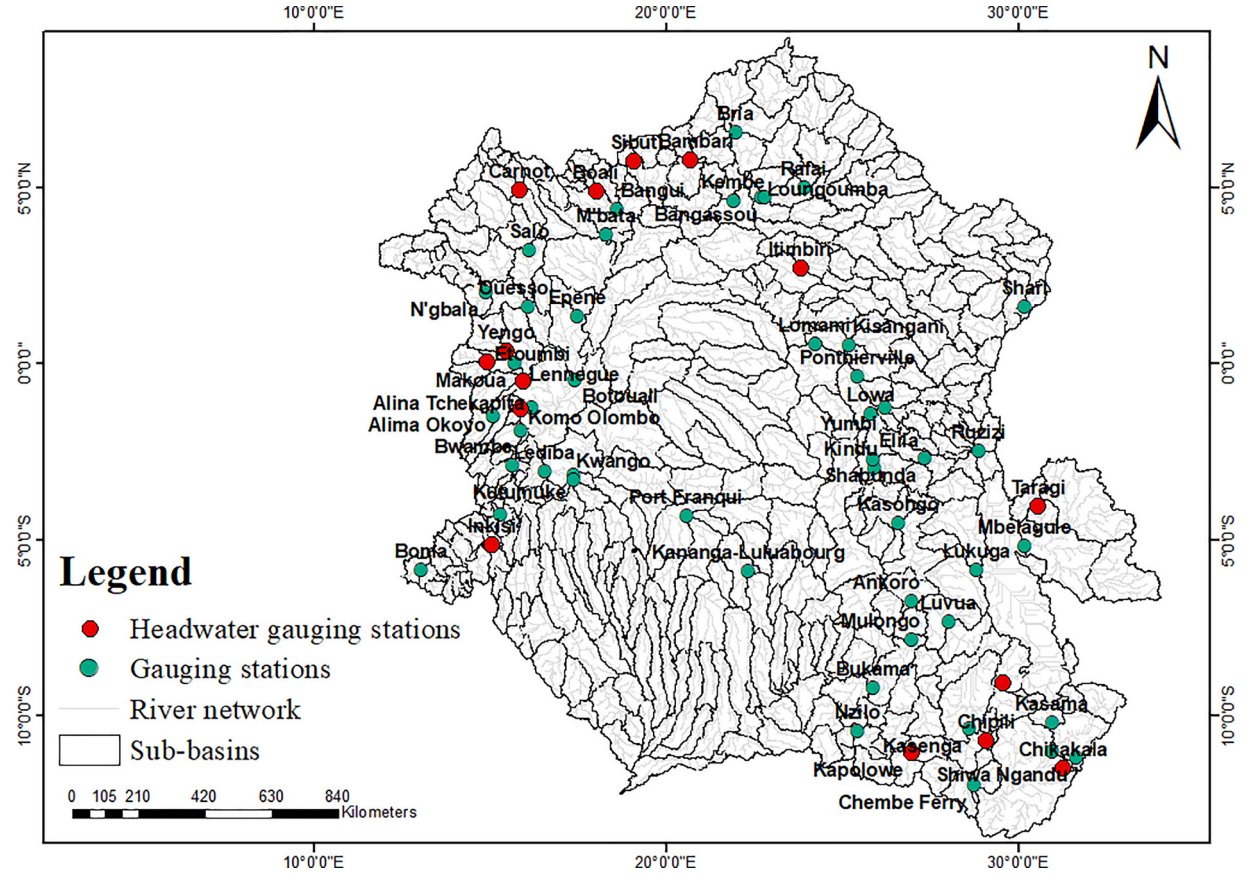

Fig. 1. Presentation of the study area showing the 403 sub-basins of the Congo River Basin, the spatial distribution of the 58 gauging stations available across the basin and the headwater gauging stations used for the quantification of hydrologic indices.

streamflow data that is available to quantify the indices. Ultimately, the indices are intended to constrain the outputs from hydrologic models and consequently reduce the uncertainty in predictions and the risks associated with water resources decision making.

2. The study area

The Congo River Basin covers a drainage area of approximately 3.7×10^6 km^2. It is the world's second largest in both size and discharge after the Amazon. In Africa, it is second only to the Nile River in length. The climate is warm and humid with two distinctive wet and dry seasons that vary with distance from the equator (Bultot, 1974; Samba et al., 2008). The mean temperature is approximately 25 °C. The mean annual rainfall is ~2 000 mm y^{-1} in the central parts of the basin, decreasing both northward and southward to ~1 100 mm y^{-1}. The annual potential evapotranspiration is between 1 100 and 1 200 mm y^{-1} and varies little across the basin (Alsdorf et al., 2016). Land cover varies from tropical evergreen forest, with little seasonal variation, in the central parts, to savannas in the north and south (Mayaux et al., 2000; Hansen et al., 2008). Similarly, the heterogeneity of the soil types and geological settings are two of the factors affecting the spatial variability in hydrologic dynamics across the basin (Tshimanga and Hughes, 2014).

The basin has four main drainage systems (Oubangui River in the north east, Sangha River in the north west, Kasai River in the south west and Lualaba River in the south east) that converge to form the main Congo River. The details of the basin topography are well documented in Runge (2007) and are not repeated here. A previous study delineated the basin into 99 modeling units, 83 resulting from an analysis of the dominant slopes and elevations, while 16 sub-basins were based on the locations of the key gauging stations (Tshimanga, 2012). As a result, the smallest modeling unit was 533 km^2, while the biggest was 185 835 km^2. In a recent study (Tshimanga et al., 2018), a similar delineation procedure was undertaken but using a revised Digital Elevation Model (MERIT DEM) (Yamazaki et al., 2017) which was corrected to remove vegetation height effects. This is particularly important for the Congo River Basin given the extent of dense tropical forest. The new delineation resulted in 403 sub-basins (Fig. 1) that are considered appropriate to represent natural hydrologic variability and that account for the current water resources management needs within the basin.

Table 1
Headwater gauging stations used for the quantification of hydrologic indices in the Congo River Basin.

SN	Station Name	Latitude	Longitude	Drainage Area		Sub-basin Code	Period of records	Months	% of missing
				Km^2	% basin				
1	Bambari	5.78	20.67	29851	0.81	O_CB355	1952-1975	282	21.27
2	Boali	4.91	18.03	4793	0.13	O_CB176	1949-1988	468	48.9
3	Carnot	4.94	15.86	18559	0.50	S_CB395	1954-1971	216	19.9
4	Chikakala	−11.49	31.28	1147	0.03	L_CB203	1970-2004	419	17.89
5	Chipili	−10.71	29.09	1321	0.04	L_CB261	1971-1981	132	0.75
6	Etoumbi	0.05	14.92	10492	0.28	S_CB236	1951-1970	240	16.6
7	Inkisi	−5.13	15.07	12824	0.35	C_CB138	1950-1959	120	22.5
8	Itimbiri	2.71	23.84	35056	0.95	C_CB185	1950-1959	120	0
9	Kapolowe	−11.04	26.95	8586	0.23	L_CB205	1933-1959	324	7.4
10	Komo Olombo	−1.28	15.87	1866	0.05	S_CB134	1963-1975	154	2.59
11	Kouyou a Linnegue	−0.50	15.93	11159	0.30	S_CB243	1953-1970	216	15.74
12	Nsama	−8.90	29.97	700	0.02	L_CB22	1959 - 2004	539	33
13	Sibut	5.73	19.08	2538	0.07	O_CB179	1951-1991	483	45.13
14	Taragi	−4.04	30.57	8395	0.23	L_CB196	1971-1979	108	5.55
15	Yengo	0.36	15.45	11686	0.32	S_CB158	1961-1980	237	21.51

2.1. Hydro-climatological data

Streamflow time series for 58 gauging stations (Fig. 1), with different periods of record, were obtained from several sources including the Global Runoff Data Centre (Fekete et al., 1999), the Office National de Recherche et du Developpement (Lempicka, 1971), Hydrosciences Montpellier—Système d'Informations Environnementales (SIEREM, http://hydrosciences.fr/sierem) and the Annuaire hydrologique du Congo Belge (Devroey, 1951-1959). Less than 25 % (15 gauging stations) of the gauging stations represent non-impacted headwater flow regimes (Fig. 1 and Table 1), while the majority are located in the downstream parts of the basin and represent cumulative streamflow characteristics from large catchment areas. Hydrologic indices obtained from headwater gauging station data are useful for establishing regional variations, but indices obtained from downstream stations will include mixtures of different upstream sub-basin hydrologic responses and will therefore be less directly useful. Gauging stations downstream of some known large wetland systems, or any major water resources infrastructure, were excluded and not used to quantify hydrologic indices.

The climate data used (rainfall and evapotranspiration) are from the Climate Research Unit (CRU TS 3.10) data for the period of 1901–2014 (Harris et al., 2014), at a spatial resolution of 0.5°. They are used to derive an aridity index (the ratio of mean annual evapotranspiration to mean annual rainfall) as a potential predictor of hydrologic behavior, as reported in many hydrologic studies (Beck et al., 2015; Tumbo and Hughes, 2015; Ndzabandzaba and Hughes, 2017; Zhang et al., 2018) and the runoff ratio. The UNIDEL (University of Delaware) rainfall dataset (covering the same period and spatial resolution) was used as to check the appropriateness of the CRU rainfall data in specific areas (Sun et al., 2018). The use of global climate datasets is justified by the lack of adequate ground-based information available for long periods and with good spatial coverage. However, the paucity of rainfall gauges over the Congo River Basin suggests that only limited observed records are used to construct and validate the global datasets, contributing to potential errors and input uncertainties (Tshimanga, 2012). While these uncertainties are likely to be quite important for hydrologic modeling, they are less likely to have a large impact on the derivation of climate indices.

2.2. Physiographic data

Physiographic data that have potential relationships with sub-basin hydrologic response characteristics are used for the classification of the sub-basins and include the topographic wetness index (TWI), slope, soil textures (fractions of silt, sand and clay) and curve number (CN). The TWI and the slope were derived from the 90 m MERIT DEM (Yamazaki et al., 2017), while soil texture data were obtained from ISRIC, the world Soil information website (Batjes, 2017). The curve number was extracted from the United States Natural Resources Conservation Service (NRCS) Runoff Curve Number (CN) dataset (Zeng et al., 2017). These physiographic properties (Table 2) have been previously reported as being important in understanding and regionalizing sub-basin runoff responses. The soil clay content has been used as a predictor for the base flow index (Beck et al., 2015), while the topographic wetness index is widely used to approximate relative soil moisture patterns (Buchanan et al., 2014) and quantify topographic control on hydrologic processes (Sørensen et al., 2006). The curve number is used in many hydrologic models (Williams et al., 2012; Zeng et al., 2017; Peña-arancibia et al., 2019; Yang et al., 2019) to simulate surface runoff generation, and is calculated based on factors such as hydrologic soil group type, land use land cover, hydrologic surface condition and antecedent moisture condition (Zeng et al., 2017).

3. Methods

The approach involves: (i) pre-processing of hydrologic data, (ii) catchment classification, (iii) quantification of hydrologic indices and development of regression relationships, and (iv) establishing of uncertainty ranges of hydrologic indices. The steps are

Table 2
Description of the climate and physiographic attributes used for the classification of the 403 sub-basins of the Congo basin.

Attribute	Unit	Description	Resolution
Climate			
AI	Unitless	The aridity index is computed as AI = PET/P, where P is the mean annual precipitation and PET is the mean annual potential evapotranspiration. Both dataset were extracted from the Climate Research Unit (CRU TS 3.10) for the period of 1901–2014 (Harris et al., 2014).	0.5°
Topography			
Slope	%	The average surface slope, computed from the MERIT DEM (Yamazaki et al., 2017)	90 m
TWI	Unitless	The average topographic wetness index, computed from the MERIT DEM (Yamazaki et al., 2017)	90 m
Land cover/Soil types			
CN	Unitless	The curve number is extracted from a global map of curve number developped by Zeng et al., 2017.	500 m
Silt	%	Soil silt content extracted from http://www.isric.org/data/AfSoilGrids (Batjes, 2017).	250 m
Clay	%	Soil clay content extracted from http://www.isric.org/data/AfSoilGrids (Batjes, 2017).	250 m
Sand	%	Soil sand content extracted from http://www.isric.org/data/AfSoilGrids (Batjes, 2017).	250 m

shown in Fig. 2.

3.1. Pre-processing of hydrologic data

The available observed streamflow time series are relatively short and have different record periods that might represent different sequences of dry and wet climatic conditions. This situation can affect the representativeness of derived hydrologic indices and therefore the data were pre-processed using a spatial interpolation approach (Hughes and Smakhtin, 1996) to extend the observed flow series to common record periods. This approach is based on the assumption that flows occurring simultaneously at sites in reasonably close proximity to each other correspond to similar percentage points on their respective duration curves. Its application requires the identification of key gauging stations within each region that have the longest record periods so that they can be used as source gauges for extending the flow series of gauging stations in their vicinity. While the method allows for the use of up to 5 source gauges with different weighting factors (Hughes and Smakhtin, 1996), only one source gauge has been used in the Congo application

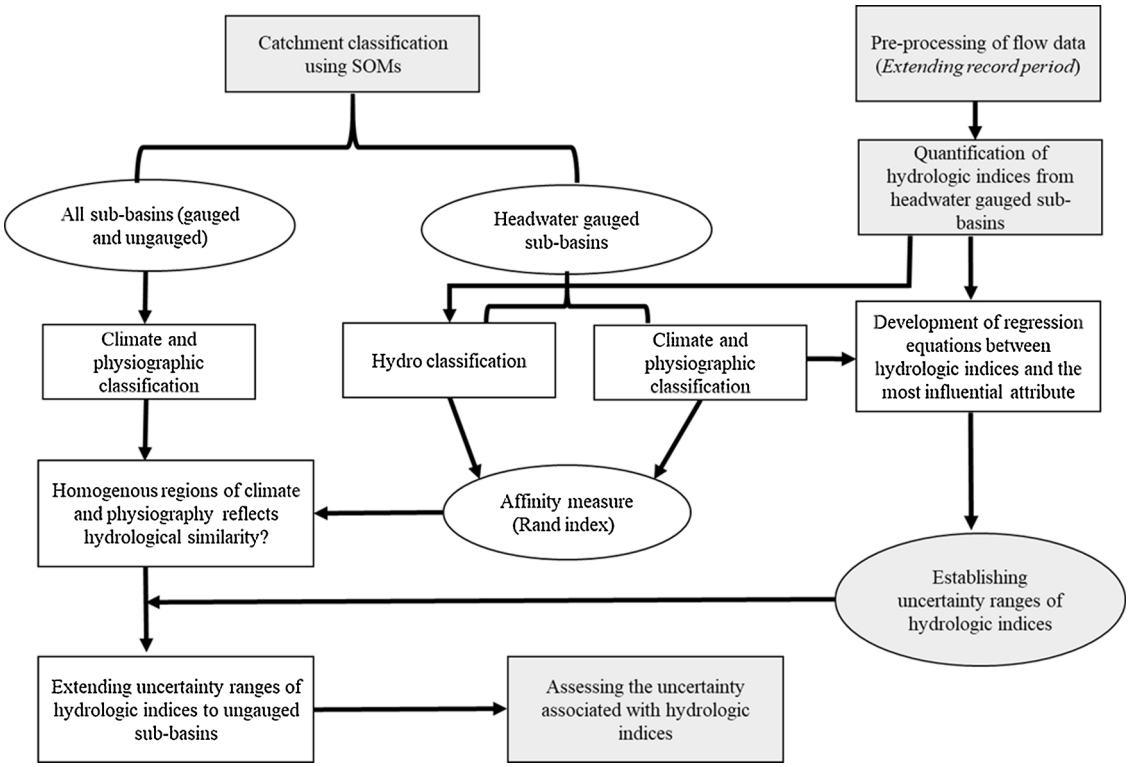

Fig. 2. Flowchart of the methodological framework including catchment classification using SOMs, hydrologic indices quantification and uncertainty ranges.

due to the limited number of available gauges. The outputs consist of both a patched (filling missing data periods) and extended time series, as well as a time series representing estimates for all months (substitute time series). The latter can then be used to compare with the original observed flow data and the reliability of the method assessed using typical objective functions (such as the Nash coefficient of efficiency).

3.2. Catchment classification

Due to the largely ungauged nature of the Congo River Basin, the available gauging stations are not sufficient to represent the variability of the hydrologic response characteristics across the different climate and physiographic regions. It is therefore necessary to assume that more readily available climate and physiographic (surrogate) information can be used in a catchment classification approach to group sub-basins into regions of expected similar hydrologic response. The limited available gauging station data can then be used to quantify the expected response characteristics of the sub-basins within the regions, albeit with a degree of uncertainty. Self-Organizing Maps (SOM) are able to analyze, organize and cluster various types of data through non-linear relationships, which represent the internal similarity of the variables. They have been used in many hydrologic applications (Hall et al., 2002; Srinivas et al., 2008; Herbst and Casper, 2008; Toth, 2009; Di Prinzio et al., 2011; Ley et al., 2011; Toth, 2013). In this study, the Viscovery SOMine software (https://www.viscovery.net/somine/) is used for the sub-basin scale similarity analysis.

The details of SOM methods are well documented in previously reported studies (Ley et al., 2011; Di Prinzio et al., 2011). A SOM consists of two layers of interconnected neurons (nodes), the input and the output layers (Kalteh et al., 2008), where the input layer represents the sub-basin attributes used in the classification, and the output layer corresponds to the number of classes to which the sub-basins are assigned. Although there are no well-defined guidelines on the appropriate number of clusters to be formed, SOM allows for both automatic (Liu et al., 2011), and user-defined cluster numbers based on the dataset and the level of detail required in the classification. For the classification to be coherent, only a few parameters need to be specified during the training cycle of a SOM. These are the map size, training parameters and clustering method. Three types of clustering methods are imbedded in the Viscovery SOMine software and include SOM-Ward, Ward and SOM-Single-Linkage. The SOM-Ward is generally used because it is considered the most efficient technique (Kohonen, 2001) and is implemented in this study. The quality of a well-trained SOM is evaluated by means of a quantization error (known also as Euclidean distance) defined as the average of the squared distance of all data records associated with a node in the output layer. It should be as small as possible and is often used as the basis for assigning input vectors (sub-basins) to nodes. Therefore, sub-basins that have similar quantization error are assigned to the same class. An independent evaluation of the accuracy of the classification is achieved with the ANOSIM statistic R that provides a test whether there exists a significant difference between the identified clusters (Clarke, 1993; Warton et al., 2012).

Prior to the SOM training, all attributes need to be standardized in order to suppress the effect of their different orders of magnitude and ensure that they all have equal importance in calculating a meaningful Euclidean distance between two points. Viscovery SOMine provides two scaling methods (Variance and Range) that work simultaneously depending on the internal distribution of each attribute. In both methods, the mean value of the attribute is subtracted from each value of the attribute so that the mean of the scaled attribute is zero. However, in the variance scaling the difference between the mean value of the attribute and each value of the attribute is divided by the standard deviation of the attribute so that the new variance of the scaled attribute will always be 1. In range scaling, the difference is multiplied by 8/(maximum-minimum) of the attribute, such that the new range is always 8. This has an advantage of reducing the impacts of outliers in the training process, thus speeding up learning and leading to faster convergence. Range scaling is automatically activated if the difference between the maximum and minimum values of an attribute is smaller than 8 times the standard deviation, otherwise the variance scaling applies.

The classification strategy adopted here was achieved in a four-step approach:

- Step (1) classifies the all 403 sub-basins based on the climate and physiographic attributes;
- Step (2) identifies, in the formed homogenous regions from step (1), all the selected headwater gauged sub-basins;
- Step (3) performs two independent classifications of the selected headwater gauged sub-basins. The first was based on the climate and physiographic attributes, and the second was based on the indices of hydrologic behavior. The hydrologic classification was performed 5 times including each index separately and all indices together in order to identify indices responsible for the highest affinity with climate and physiographic attributes;
- Step (4) compares the two independent classifications obtained in step (3) to assess the level of overlap. This was done using an index of affinity developed by Rand (1971).

The rand affinity index (Rand, 1971; Di Prinzio et al., 2011; Ssegane et al., 2012) was calculated using the following expression:

$$R = \frac{a + b}{a + b + c + d} \tag{1}$$

Where R varies between 1 (perfect agreement between the two pools of clusters) and 0 (no agreement). The meaning of the terms a, b, c and d is given under the following assumptions:

Consider two classifications (C1 and C2) of the same dataset, a pair of sub-basins can be assigned to the same class or different clusters in C1 and C2. So, “a” is defined as the number of sub-basin pairs that are in the same cluster in classification C1 and in the same cluster in classification C2; “b” as the number of sub-basin pairs that are in different clusters in C1 and in different in C2; “c” as the number of sub-basin pairs that are in the same cluster in C1 but different clusters in C2; “d” as the number of sub-basin pairs that

are in different clusters in C1 but in the same cluster in C2.

3.3. Derivation of the indices and their relationships

The choice of hydrologic indices depends on the objectives of the study, which in this case are focused on establishing indices that are suitable for constraining the outputs of a specific rainfall-runoff model (the Pitman model: Tumbo and Hughes, 2015; Ndzabandzaba and Hughes, 2017). The hydrologic indices used in the current version of this model (Pitman model) are the mean monthly runoff volume (MMQ in $m^3 * 10^6$), the mean monthly groundwater recharge (MMR in mm), the 10th, 50th and 90th percentiles of the flow duration curve (FDC) expressed as a fraction of MMQ and the percentage of zero flow. However, this study is limited to the mean monthly runoff volume and the 10th, 50th and 90th percentiles of the flow duration curve (FDC) because they can be directly estimated from the available streamflow data and zero flows are not relevant to the Congo River Basin at the scale of the sub-basins used in the study. We could have also used 33rd and 66th, which are sometimes used to compute the slope of the flow duration curve. However, the 10th, 50th and 90th percentiles are considered here to represent the minimum number of key indices that can characterise the complete flow duration curve, for they consider more extreme high and low flows. Essentially, we want to capture a wider range of flow behaviour than the 33rd and 66th percentiles represent. These percentiles are also used within the modelling software that will be used.

The mean monthly runoff volume (MMQ) was expressed as runoff ratio (RR) in order to suppress sub-basin scale effects:

$$RR = \frac{MMQ}{P} \tag{2}$$

Where, MMQ is the long-term average monthly streamflow and P the long-term average monthly precipitation. The RR represents the long-term water balance separation between water being released from the sub-basin as streamflow and as evapotranspiration.

The three percentiles of the flow duration curve represent the frequency distribution of flows of different magnitude, the 10th percentile represent high flows, the 50th medium flows and the 90th low flow conditions.

The classification approach used in this study assumes that streamflow indices exhibit some consistent relationships with sub-basin climate and physiographic characteristics (Yadav et al., 2007). However, this study also follows the approach applied by Tumbo and Hughes (2015) and Ndzabandzaba and Hughes (2017) where the relationships between the same hydrologic indices and the aridity index are explored. It is assumed that any relationships used to estimate hydrologic indices in un-gauged sub-basins will necessarily be uncertain and therefore 90 % regression confidence limits are used to quantify the degree of uncertainty.

3.4. Spatial disaggregation of flow time series

Due to the low number of gauging stations, the relationships between the climate/physiographic attributes and the hydrologic indices can result in high uncertainty. However, there are several gauging stations that include a relatively small number of sub-basins (≤ 6) that can be included if an appropriate method of spatially disaggregating the total downstream response characteristics can be used. While there may be several alternative approaches, the current study adopted an iterative process. Two gauging stations (in the north of the Congo River Basin), which represent eleven sub-basins, were selected and used to derive estimates of the sub-basin hydrologic indices. None of these have any identified upstream anthropogenic impacts and are not influenced by wetland effects. The iterative process essentially involves using the initial regression relationships (developed from the gauged headwater sub-basins) to provide initial estimates for the 11 additional sub-basins. This step ensures that sub-basin relative differences in response are consistent with the initial relationships. The flow percentile indices are converted to absolute values (i.e. not as fractions of MMQ) and are summed to give cumulative values at the gauging station. All of the cumulative index values are compared to the observed gauging station values and correction factors determined for each index. These correction factors are then applied to the sub-basin initial estimates and the FDC indices converted back to fractional values by dividing by MMQ. Clearly, the indices for the additional 11 data points are less certain than those derived from the gauged headwaters. However, the approach is justified on the basis of the very limited number of gauged headwater sub-basins and at least allows some of the response characteristics of the two larger gauged catchments to be included in a second round of regression analysis.

3.5. Validating the uncertainty ranges and assessing the uncertainty

Two types of independent information on hydrologic indices were used to validate the derived ranges of indices. The first source is made up of published runoff ratios for specific areas across the basin (Snel, 1957; Laraque et al., 1998). The aridity index values of these areas were used to check whether the runoff ratio could fit within the computed ranges. The second source of information is based on the cumulative flow time series at downstream gauged stations that have no substantial attenuation effects. The hydrologic indices were computed at these gauges and plotted against the area-weighted values of the sub-basin aridity indices. One of the major sources of uncertainty in the quantified hydrologic indices involve the input rainfall data. It has been acknowledged that the reliability of a rainfall dataset is mainly limited by the number and the spatial coverage of surface stations (Sun et al., 2018) and the CRU rainfall dataset used in this study is based on a very limited amount of observed rainfall data. Previous studies reported on the lack of agreement between different interpolated rainfall datasets (Sun et al., 2018). The consistency of the CRU and the UNIDEL (Sun et al., 2018) datasets is checked by computing ratios of UNIDEL to CRU mean annual rainfall. Regions where the computed ratio is high (e.g. > 1.2) were identified as potential areas of high uncertainty where the uncertainty ranges of the indices would need further

Table 3
Extended flow record periods of gauging stations used for the derivation of hydrologic indices in the Congo River Basin.

Gauging station	Sub-basin code	Original record period	Extended record period	Source gauging station
Bambari	O_CB355	1952 - 1975	1949 - 1988	O_CB176
Boali*	O_CB176	1949-1988	Not available	Not available
Bria@	O_CB95	1954 - 1978	1949 - 1988	O_CB176
Carnot	S_CB395	1954 - 1971	1949 - 1988	O_CB176
Chikakala*	L_CB203	1970-2004	Not available	Not available
Chipili	L_CB261	1971 -1981	1970 - 2004	L_CB203
Etoumbi	S_CB236	1951 - 1970	1951 - 1980	S_CB158
Inkisi#	C_CB138	1950-1959	Not extended	Not available
Itimbiri	C_CB185	1950-1959	1949 - 1988	O_CB176
Kapolowe	L_CB205	1933 - 1959	1921 - 1959	L_CB27
Komo Olombo	S_CB134	1963 - 1975	1951 - 1980	S_CB158
Kouyou a Linnegue	S_CB243	1953 - 1970	1951 - 1980	S_CB158
Rafai@	O_CB181	1952 - 1973	1949 - 1988	O_CB176
Sibut	O_CB179	1951 - 1991	1949 - 1988	O_CB176
Taragi	L_CB196	1971 - 1979	1970 - 2004	L_CB203
Yengo	S_CB158	1961 - 1980	1951 - 1980	S_CB236

* Gauging stations having long period of records with no need of extension.
\# gauging station with record period not extended because of lack of donor gauge in its vicinity.
@ New gauging stations added for the spatial disaggregation procedure (see Section 3.4).

refinement.

4. Results

4.1. Extended streamflow series

It was important to obtain common record periods of streamflow time series in order to ensure the representativeness of the computed hydrologic indices and minimize the differential effects of the number of wet and dry periods represented in short record periods. Table 3 lists the original and extended record periods, as well as the source gauging stations used for the record extension. Fig. 3a shows a plot of two streamflow series at one gauging station (L_CB261) located in the upper Lualaba sub-basin and illustrates the reliability of the approach by comparing the observed flows with the substitute time series (i.e. all months estimated from the source gauge). The general pattern of the observed streamflow series is well reproduced and therefore the extended and infilled records should be adequately reliable. Similar results were obtained for all the gauging stations and their Nash coefficients of efficiency, based on comparisons between the observed and substitute flows are shown in Fig. 3b. For the majority of gauging stations the Nash coefficient of efficiency is above 0.6 for both low and high flows. The final hydrologic indices were derived from the patched streamflow series (i.e. a combination of observed, patched and extended data).

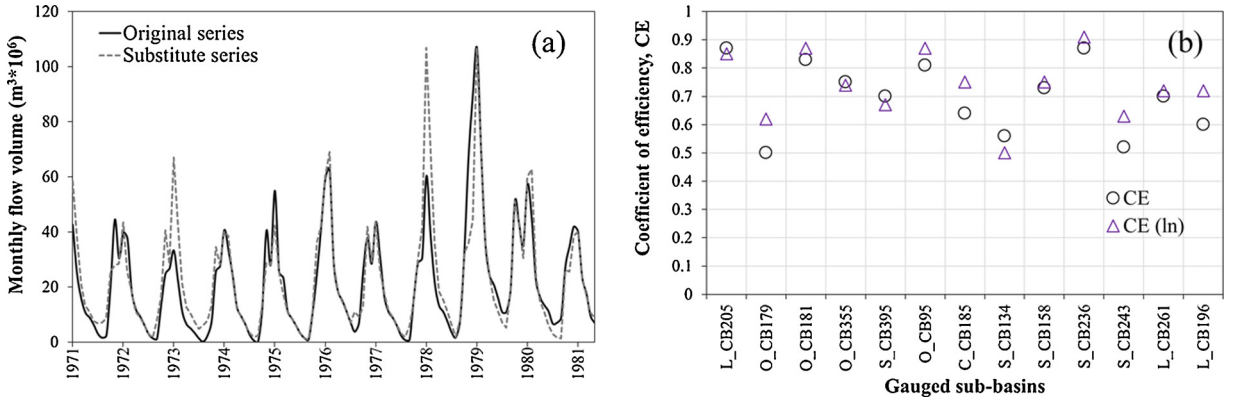

Fig. 3. Results of the pre-processing analysis of flow data. (a) graphical comparison between the original flow series and the substitute series (at L_CB261 gauging station). The latter is defined as estimated values of flow records that overlap with the original series. (b) Nash coefficients of efficiency for low (CE ln) and high (CE) flows showing how well the original series are reproduced during the extension procedure of the flow record period.

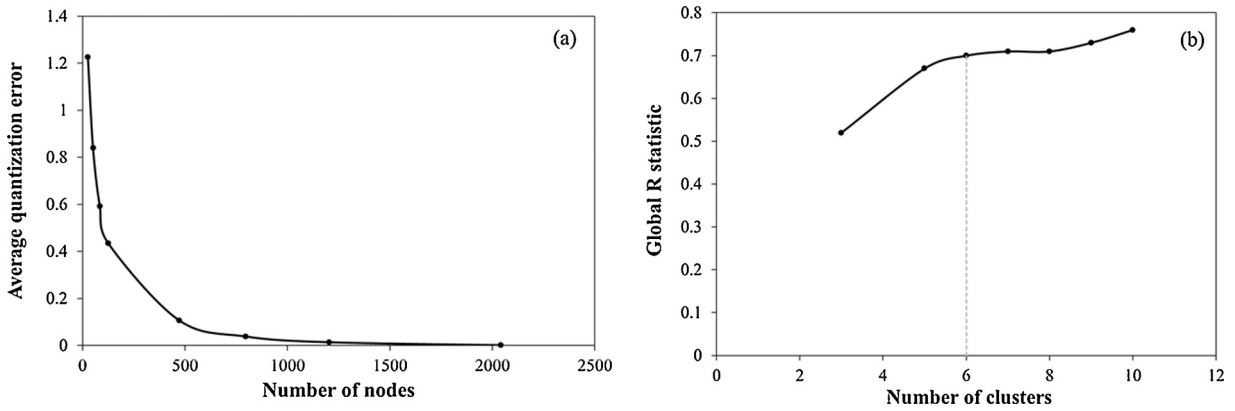

Fig. 4. Optimal size of the trained maps through SOM and the selection of optimal number of clusters. (a) asymptotic decrease of the quantization error showing the lowest achieved with 2000 nodes. (b) the evolution of the ANOSIM statistic R with the number of clusters. The dashed line shows the optimal number of clusters with a global R statistic of 0.7.

4.2. Classification by climate and physiographic characteristics

The classification of the 403 sub-basins of the Congo River Basin was achieved by training a self-organizing map (SOM). Fig. 4a. displays different levels of the quantization error (QE) obtained from using different map sizes. The results show that the larger the map size (number of nodes), the smaller the quantization error, which expresses the adequateness of representing the input vectors by a specific node. Out of seven attributes used for the classification, only five were retained, while the TWI and Sand attributes were removed because they were highly correlated with Slope and Clay, respectively. Therefore, a map size of 2 000 nodes (QE = 0.0003) was judged appropriate in representing the dataset and was used for the classification. Different numbers of clusters were also tested in order to obtain the number that maximize the within group similarity and between group dissimilarity. The application of an independent quality measure (ANOSIM statistic R: Fig. 4b) suggests that six clusters are appropriate for the Congo River Basin. These homogenous groups are significantly different with a global R of 0.7 at a p-value of 0.001 %. The R statistic increases with an increase in the number of clusters, but this increase is associated with low dissimilarity ($0.25 < R < 0.48$) between some clusters when greater than six clusters are used. In contrast, with six clusters all the between-group dissimilarity values are above 0.53. Table 4 displays the dissimilarity matrix of the six homogeneous groups where high values of R implies high dissimilarity between groups. Overall, the average squared distances within groups were far smaller than the average squared distances between groups.

The spatial distribution of the six climate and physiographic regions resulting from the classification is shown in Fig. 5. The groups are coherent and preserve a high degree of spatial proximity. Approximately 28 % of sub-basins are assigned to Region 1, 19.4 % to Region 2 and 3, 14.6 % to Region 4, 11.6 % to Region 5 and 7.2 % to Region 6. Fig. 6 illustrates the spatial relationships between the climate and physiographic attributes across the six regions, while Table 5 provides descriptions of each homogenous region. Region 1 has the highest number of gauged headwater sub-basins, while Region 6 has none.

4.3. Classification by hydrologic behavior

Based on the distribution of gauged sub-basins, two independent classifications, each having five groups, were performed using a similar approach as highlighted in Section 3.2. The comparison between the physiographic classification and the five classifications based on hydrologic behavior is shown in Fig. 7. Only the three fractions of the FDC indices and runoff ratio were used to represent sub-basin response behavior, while five climate and physiographic attributes represented sub-basins physical properties. Overall, there exists a high degree of affinity (Rand index = 73 %) between the physiographic classification and the hydrologic classification when all indices of hydrologic behavior are used in the classification. However, the highest affinity (Rand index = 82 %) is achieved when only the Q50/MMQ index is used. This suggests that the climate and physiographic attributes used in this classification are able

Table 4
Dissimilarity matrix (ANOSIM statistic R) of the six climate and physiographic regions showing that all regions are significantly different at p = 0.001 %

	Region 1	Region 2	Region 3	Region 4	Region 5
Region 1					
Region 2	0.53				
Region 3	0.53	0.52			
Region 4	0.89	0.69	0.85		
Region 5	0.85	0.85	0.73	0.71	
Region 6	0.62	0.86	0.83	0.88	0.52

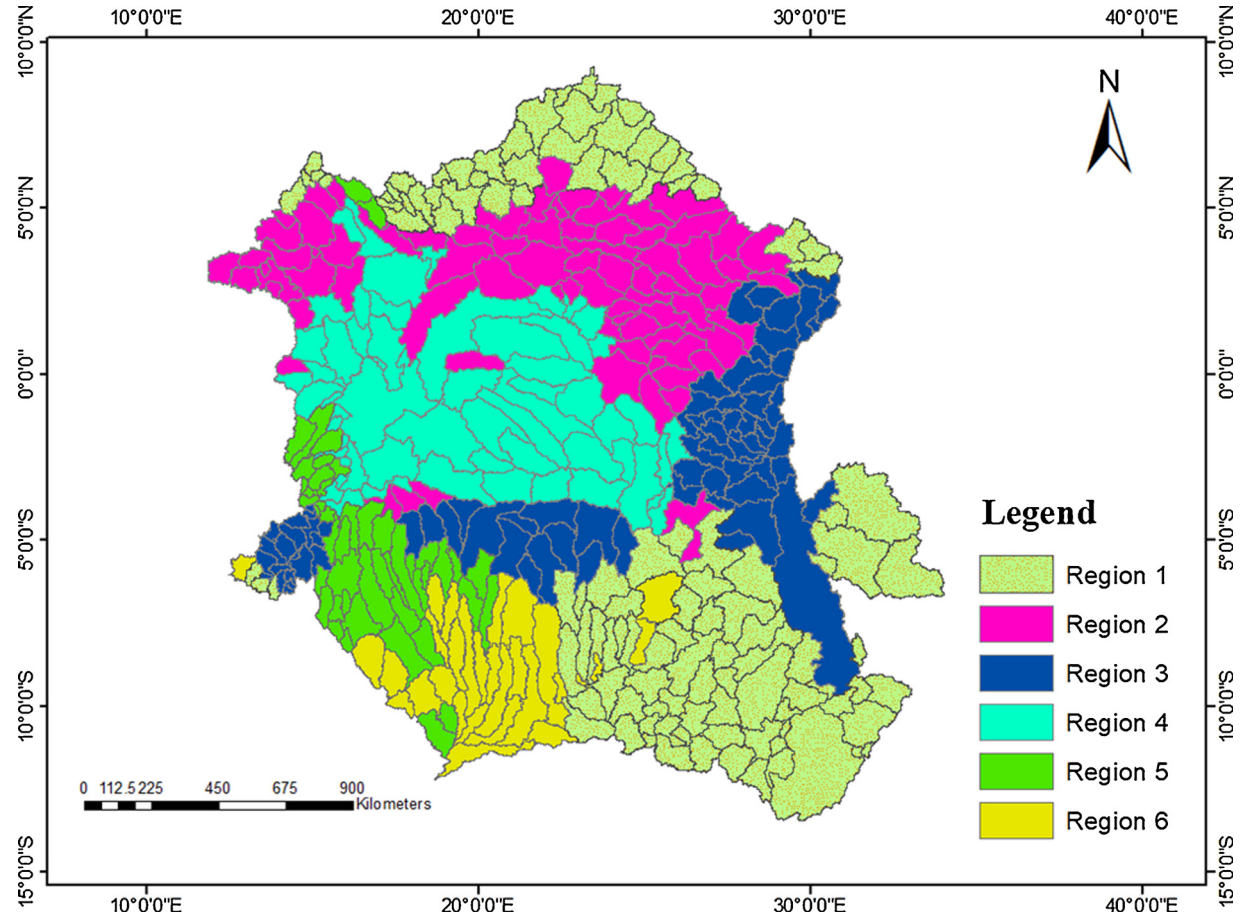

Fig. 5. Spatial distribution of the six homogenous regions of sub-basins of similar climate and physiographic properties in the Congo River Basin identified from the application of SOM.

to capture different components of the sub-basin's hydrologic response at different degrees. Due to the limited number of gauged headwater sub-basins, some clusters are made up of less than 3 sub-basins (Table S1 in supplementary information). The implication of this result is that predictive equations for the hydrologic indices in most of the clusters cannot be developed, due to a lack of enough sample points. For those clusters where there are 4 representative gauges, Table 6 illustrates that there are potentially strong relationships between the physiographic/climate attributes and the hydrologic indices. However, the shape of the relationships is regionally variable and because of a lack of enough gauging stations to represent some clusters (Fig. 8), we cannot extrapolate the relationships to all the sub-basins in the Congo. In contrast, when all the sub-basins are included in the regression analysis, the aridity index is revealed as the best single predictor of hydrologic behavior (Table 7), with CN second. The other attributes do not appear to offer any additional predictive value (Table 7).

4.4. Aridity index as a control on hydrologic indices

The spatial pattern of the aridity index (Fig. 9) across the Congo River Basin exhibits some degree of similarity with the six homogenous regions (Fig. 5). A concentric pattern of aridity index suggests the increase of the aridity index from the Cuvette Centrale towards headwater tributaries located in the north, east and south parts of the basin.

Fig. 10a shows the plot of the runoff ratio (RR) as a function of arity index (AI). Overall, a high aridity index results in a low runoff ratio and vice versa. For instance, the majority of gauging stations representing sub-basins located in the southeast and north of the Congo River Basin (Region 1) show high values of aridity index, indicating a low runoff ratio (RR $<$ 0.25) potential. The observed pattern in humid conditions (AI $<$ 0.75) does not show a similar pattern of RR as in Region 1. With very limited information in this region 4, we cannot confirm the apparent increase of RR with the increase of AI. In general, while two discernible relationships could have been established between the AI and the runoff ratio, it would have been difficult to determine the basis on which ungauged sub-basins could have been assigned to either relationship, given the fact that observed gauging stations of the same region would be on different regression lines. Therefore, a single regression relationship (R^2 = 0.63) is derived between the aridity index and the runoff ratio.

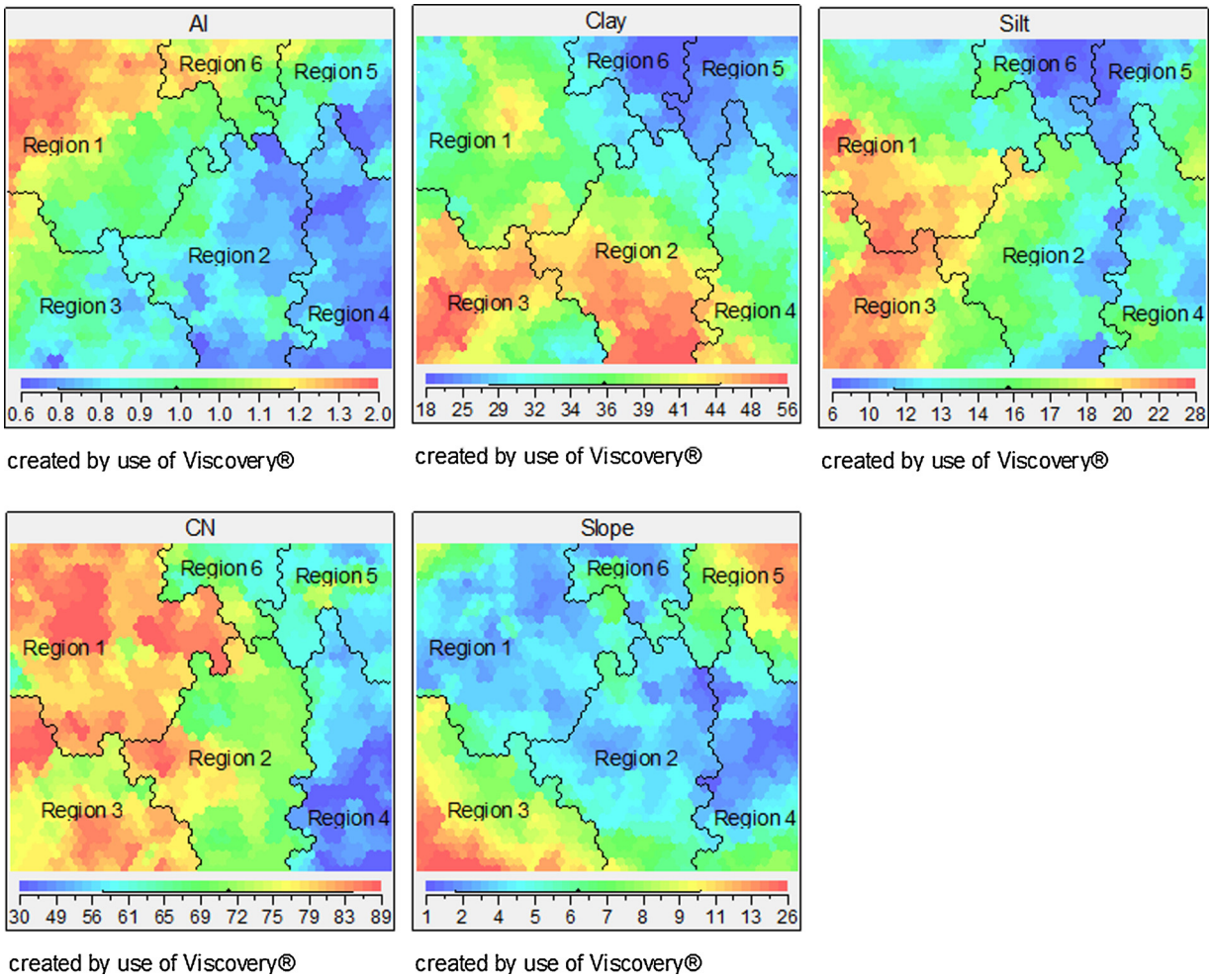

Fig. 6. Non-linear relationships between climate and physiographic attributes (Aridity index, Clay, Silt, Curve number and Slope) across the six regions of the Congo River Basin obtained from the application of SOM. The red colour represents high values of the attributes and blue low values.

A positive power relationship (R^2 = 0.84) between the aridity index and the Q10/MMQ is shown in Fig. 10b. Sub-basins of Region 4 are located at the bottom end of the regression line, indicating a slow to moderate response to rainfall inputs. In contrast, sub-basins of Region 1 are spread throughout the regression line. The Q50/MMQ index (Fig. 10c) shows a positive power relationship (R^2 = 0.81) with the aridity index and follows a similar trend to the runoff ratio, where high values of aridity index were associated with low indices. The Q90/MMQ index also exhibits a positive power relationship (Fig. 10d) but with a higher degree of scatter (R^2 = 0.725). Comparing Fig. 10b and d, suggests that Regions 2 and 4 have regimes with low variability, while Region 3 is more variable and Region 1 is represented by flow regimes of different degrees of variability. Without more data points it is difficult to predict if areas with higher aridity indices would have very low Q90/MMQ indices and possibly zero flows for some of the time. This could depend on the spatial scale of the sub-basins, in that aggregation of contributions from different parts of a large sub-basin, coupled with the effects of flow routing, suggest that zero flow conditions are unlikely. However, small sub-basins might experience zero flows, but the majority of the sub-basins used in the current modeling units (Tshimanga et al., 2018) are greater than 2 000 km^2, limiting the possibility of zero flow at a monthly time scale.

Through the application of the spatial disaggregation procedure (Section 3.4), estimates of the hydrologic indice values are available for an additional 11 sub-basins (Table S2 in supplementary information). While these values are more uncertain than those based on the gauged headwater sub-basins, they have been included to expand the data set before calculating regression line confidence intervals to quantify the uncertainty ranges of the hydrologic indices.

4.5. Uncertainty ranges of hydrologic indices

Fig. 11 shows the 90 % confidence intervals based on the updated regression relationships after the inclusion of the data for additional 11 sub-basins disaggregated from downstream gauging station data. A comparion between Figs. 10 and 11 suggests that the shape of the relationships has hardly changed, while most of the R^2 values have slightly decreased and therefore the final

Table 5
Description of the six homogenous regions of the Congo River Basin obtained from the application of SOM.

Region	Number of sub-basins per region	Number of gauged sub-basins per region	Description
1	112	8	Characterized by high values of aridity index (AI), medium content of clay and flat to undulating topography (slope). These three attributes account for more than 75 % to the within group similarity. Sub-basins are mostly located in the south-eastern and northern parts of the Congo River Basin. The presence of high values of curve number suggests the potential of the sub-basins to be dominated by surface hydrological processes, making them prone to flash flooding.
2	78	1	Dominated by soil texture with high content of clay resulting in a decrease of silt content with medium to high values of curve number. These attributes contribute by more than 85 % to the overall similarity within the region. These conditions suggest the potential of the region to limited infiltration rate while maintaining appropriate level of humidity (AI $<$ 1) on flat to undulating topography. The majority of the sub-basins are located in the north-eastern part of the Cuvette central, while the others are specifically located in the Sangha drainage system.
3	78	1	Mostly dominated by high clay content, high slope and low aridity, thus accounting for more than 85 % of the within region similarity. In contrast to region 2, this region is characterized by high silt content and high slope, suggesting the dominance of sub-surface processes particularly interflow. The sub-basins are mostly located in the eastern mountainous region of the Congo, known as rift valley. Similar conditions are found in the south of the Cuvette central and the lower Congo River before exiting to the Atlantic Ocean.
4	59	4	Represents most of sub-basins located in the Cuvette central, the central part of the Congo River Basin. More than 80 % of within group similarity is controlled by CN, Slope and Silt. Low values of curve number suggest high infiltration rate resulting in predominance of sub-surface processes over the surface processes. The climate is humid with lowest values of aridity index and flat to undulating topography. These conditions portray the prevalence of the accumulation processes of the eroded materials coming from all the upstream tributaries, thus favouring factors that contribute to the formation of the wetlands and channels with high degree of sinuosity and braiding.
5	47	1	Clay, Slope and Silt represent more than 70 % of within group similarity. The soils characteristics (low clay and silt) imply the dominance of the infiltration processes, resulting in high storage capacity. Sub-basins of the Batéké plateau system, located in western part of the Congo River Basin, are found in this group and are mostly characterized by v-shaped valleys with deep soils, suggesting the presence of groundwater aquifer systems with high storage capacity. However, a humid climate (AI $<$ 1) on undulating to steep topography characterizes this region.
6	29	0	Almost similar characteristics as region 5, but the difference resides in that region 6 has arid climate (0.94 $<$ AI $<$ 1.3) and flat to undulating topography. The sub-basins are located in southern part of the Kasai drainage system. Clay, Silt and AI account for more than 75 % to the within group similarity.

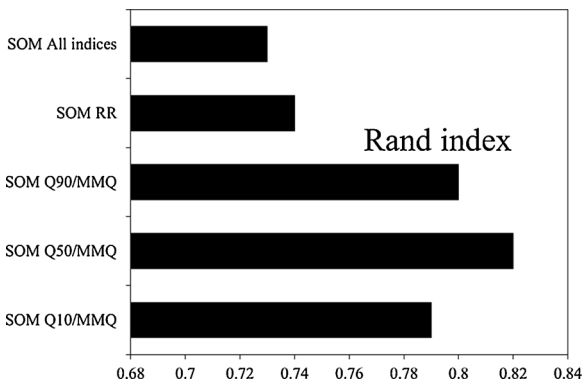

Fig. 7. Bar-diagram for Rand index showing the comparison between hydrologic and physiographic classifications. “SOM all indices”, “SOM RR”, “SOM Q90/MMQ”, “SOM Q50/MMQ” and “SOM Q10/MMQ” are independent SOM based classifications of the 15 gauged headwater sub-basins based only on all the indices, runoff ratio, Q90/MMQ, Q50/MMQ and Q10/MMQ, respectively. All these hydrologic classifications are compared with the climate and physiographic classification of the 15 gauged headwater sub-basins.

uncertainty bounds increased. The degree of uncertainty varies according to the type of the hydrologic index, with greater relative uncertainty in the runoff ratio and the Q90/MMQ indices, while relatively low uncertainty characterizes the Q10/MMQ and Q50/MMQ indices.

Given that the available data to quantify the hydrologic indices is very limited, it is impossible to make any firm a priori

Table 6
Highest R^2 of the relationship observed between physiographic attributes and hydrologic indices within clusters/groups having 4 sub-basins formed from 15 gauged headwater. In bracket are correlation coefficients and in bold the highest R^2 for each hydrologic index.

	AI	Slope	CN	Silt	Clay
RR	0.96 (-0.97)	0.59 (0.74)	0.64 (0.76)	0.96 (-0.98)	**0.98 (-0.99)**
Q10/MMQ	**0.97 (0.99)**	0.94 (0.98)	0.79 (0.92)	0.67 (0.84)	0.35 (-0.57)
Q50/MMQ	0.75 (-0.84)	0.63 (-0.78)	**0.95 (-0.99)**	0.64 (-0.79)	0.44 (-0.65)
Q90/MMQ	0.46 (-0.68)	0.46 (-0.68)	0.08 (-0.34)	0.55 (-0.73)	**0.97 (0.98)**

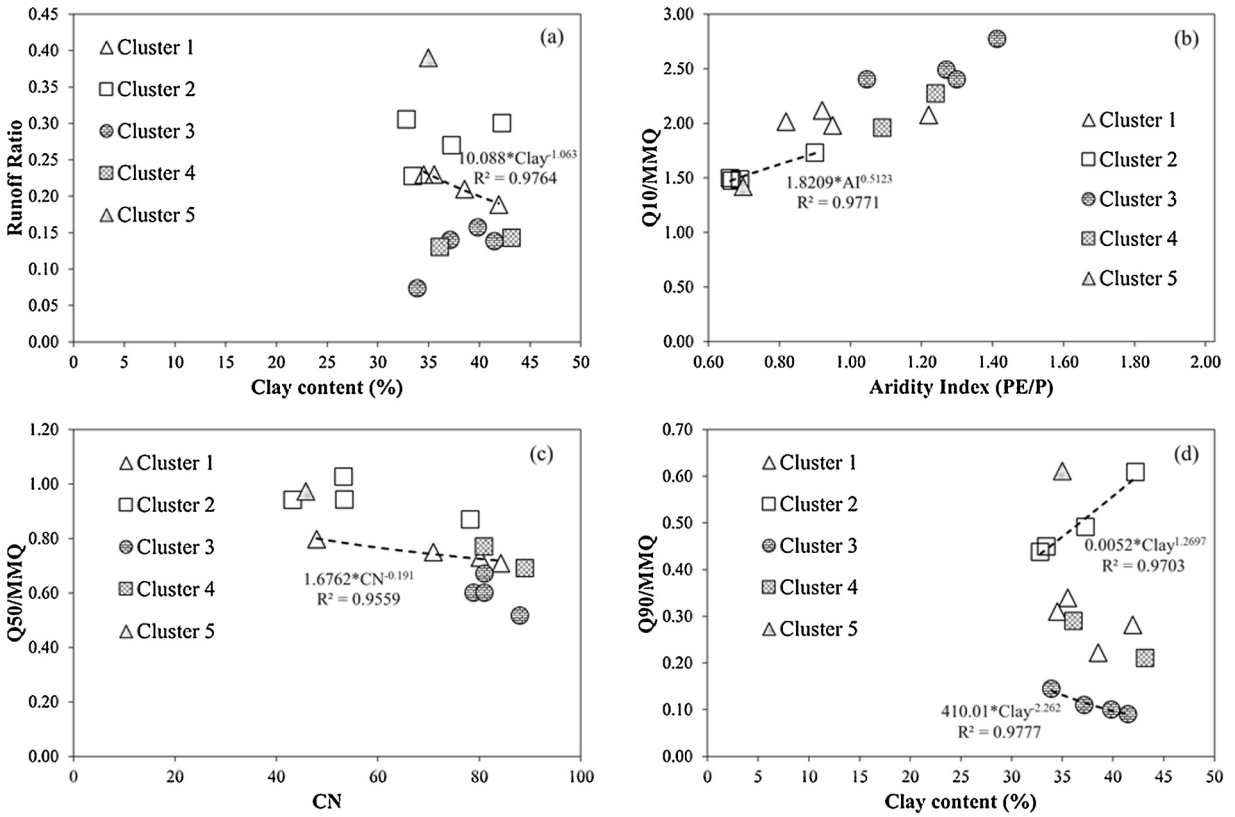

Fig. 8. Potential regression relationships between the climate/physiographic attributes and the hydrologic indices within clusters formed from the 15 gauged headwater sub-basins. (a) Runoff ratio seems to develop a relationship with Clay in cluster 1, (b) Q10/MMQ index seems to have a relationship with AI in cluster 2, (c) Q50/MMQ index seems to have a relationship with Curve number in cluster 1 and (d) Q90/MMQ index seems to develop relationships with clay in clusters 2 and 3.

Table 7
Coefficient of determination of power regression relationship between hydrologic and physiographic attributes across 15 gauging stations. AI and CN are potential predictors.

	AI	Slope	CN	Silt	Clay
RR	0.62	0.26	0.35	0.062	0.15
Q10/MMQ	0.84	0.12	0.61	0.11	0.12
Q50/MMQ	0.81	0.16	0.59	0.0006	0.13
Q90/MMQ	0.72	0.29	0.24	0.057	0.18

statements about how representative these uncertainty bound relationships are for all of the sub-basins of the Congo River Basin. However, Ndzabandzaba and Hughes (2017) proposed that checks can only be made by using the uncertainty bounds to constrain the individual sub-basin responses in a hydrological model and assess the outputs against observed data representing aggregated responses downstream. Some limited independent estimates of runoff ratio (Fig. 12) suggest that many of the reported runoff ratios fit within the computed bounds. Notable exceptions are some estimates for the Batéké plateau system in the western Congo River Basin (Region 5), and for the rift valley region (eastern Congo River Basin). There are quite large uncertainties in the rainfall data for these

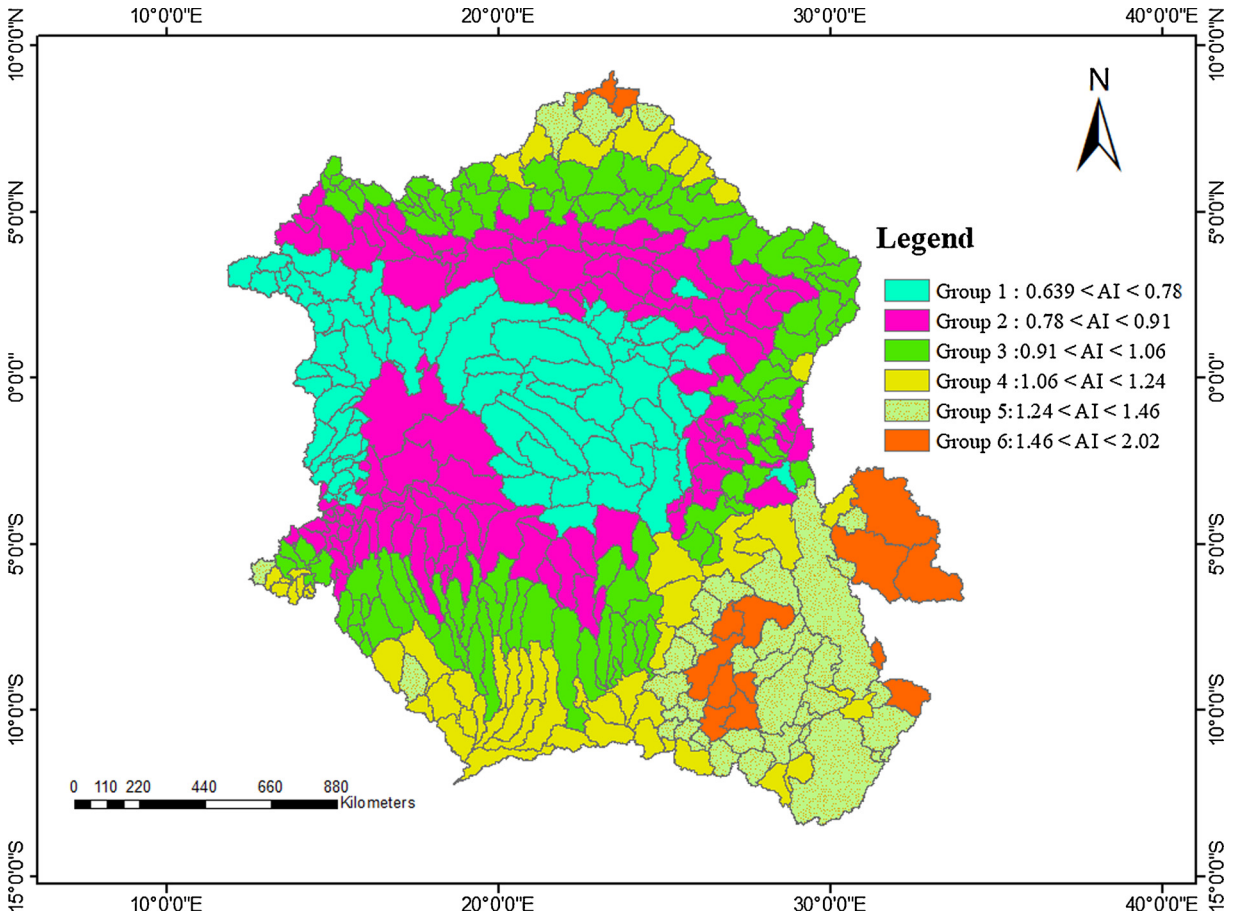

Fig. 9. Spatial pattern of the aridity index (PE/P) across the Congo River Basin showing a concentric pattern. The aridity index increases from the Cuvette Centrale towards north, east and south headwater tributaries of the Congo River Basin.

steep areas, but even if a different rainfall data set is used (UNIDEL), the estimated runoff ratios remain well outside the computed uncertainty bounds. In contrast, the independent estimates in flat to undulating topography (0.5–5%) regions (Cuvette Centrale, Northeast Congo, upper Lualaba and Southeast Kasai), generally fall within the uncertainty bounds regardless of which rainfall data set is used. The average runoff ratio (0.24) for the whole Congo River Basin (Laraque and Olivry, 1996) also falls within the uncertainty bounds.

5. Discussion and conclusion

It is a common practice in hydrology to use catchment classification as a means of extending hydrologic information from gauged to ungauged sub-basins. This procedure requires each region to have a predictive equation of hydrologic response based on potential climate and physiographic predictors (Yadav et al., 2007; Kapangaziwiri et al., 2012). The success of this approach largely depends on the number of gauged sub-basins with sufficiently long records. The classification of the 403 sub-basins of the Congo River (Fig. 5 and Table 5) demonstrated that the climate and physiographic attributes used in this study can identify relatively homogeneous regions, suggesting that there is a potential to interpret hydrologic similarity based on similarity in climate and physiography (Oudin et al., 2010; Ley et al., 2011). Furthermore, the hydrologic classification based on the 15 gauged sub-basins showed that the aridity index, the surface slope, the curve number, the silt and clay contents are potential predictors of hydrologic response indices. However, the number of gauged sub-basins (even including the disaggregated data for a few larger gauged sub-basins) is not sufficient to develop individual predictive relationships between the hydrologic indices and the climate and/or physiographic attributes for each of the identified regions. Some of the clusters (identified in the hydrologic classification: Section 4.3) are represented by less than 3 sample gauged sub-basins. This is part of the challenge of data scarcity in many parts of the world, including the Congo River Basin. Fortunately, the alternative approach of developing generic predictive equations for the whole basin generated results that produce acceptable levels of uncertainty, as measured by the width of the confidence intervals around the regression relationships. While several approaches to developing these equations were assessed using combinations of the climate and physiographic attributes (informed by the results of the hydrologic classification), it transpired that the aridity index produced the best results for all of the

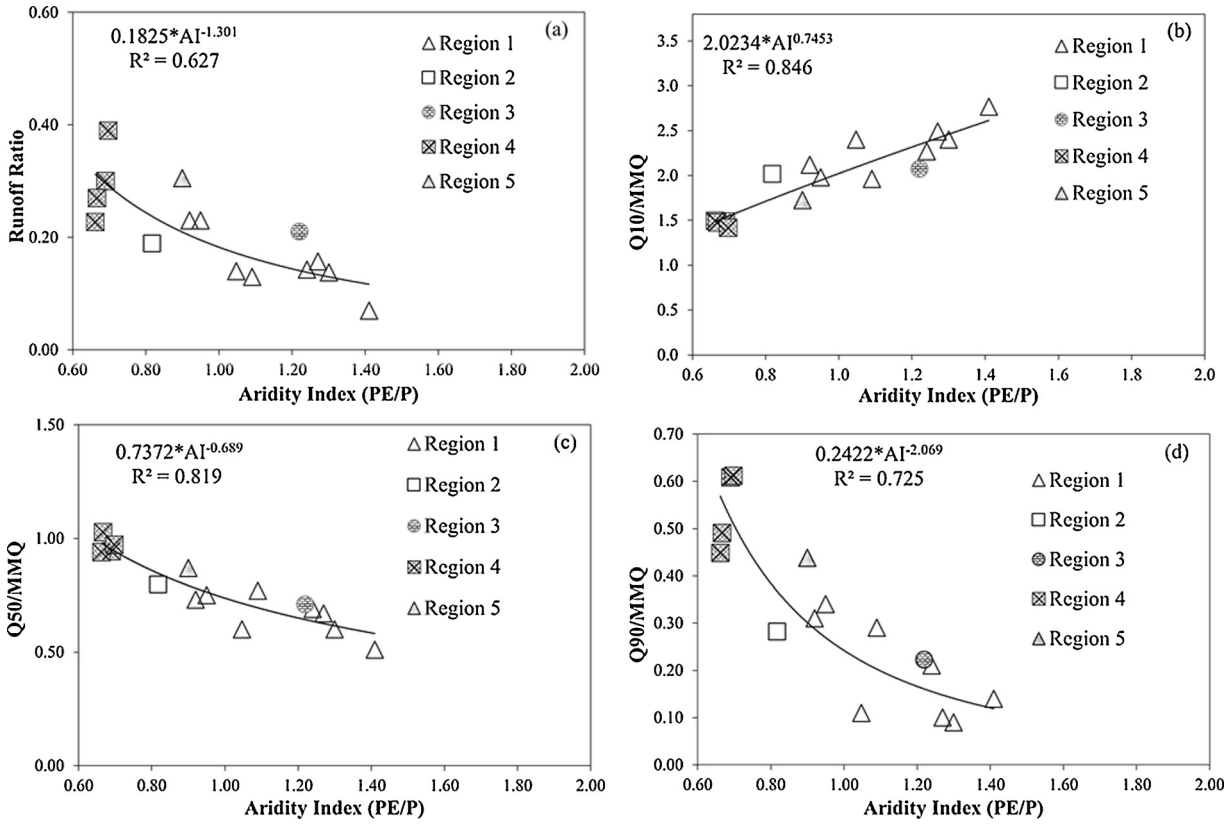

Fig. 10. Power regression relationships between the aridity index and the hydrologic indices across headwater gauged sub-basins of the Congo River Basin. (a) Runoff ratio, (b) Q10/MMQ index, (c) Q50/MMQ index and (d) Q90/MMQ index. Regions refer to those obtained through the application of SOM to all 403 sub-basins of the Congo based on climate and physiographic attributes.

hydrologic indices.

A comparison between the spatial patterns of climate/physiographic regions (Fig. 5) and the aridity index groups (Fig. 9) shows some distinct similarities. The areas where there are fewer similarities are mainly within or across those regions that are less distinguishable (based on the ANOSIM statistics in Table 4) from other regions. Regions 1, 2 and 3 have low ANOSIM statistics and also cover a wide range of aridity groups (2–6), while Regions 4 and 5 have the highest ANOSIM statistics and all the sub-basins have generally low aridity values. Similarly, Region 6 is generally distinguishable from the other regions and most sub-areas fall within aridity groups 3 and 4. The conclusion is that, although the original climate/physiographic regionalization results are not used in the developed predictive equations for the hydrologic indices, the spatial patterns of variability in hydrologic response are not too dissimilar to the originally identified regions.

The fact that the aridity index emerges as the best available predictor of hydrologic indices (Table 7) in the Congo River Basin is perhaps not surprising, as the same index has been used successfully in other parts of southern Africa (Tumbo and Hughes, 2015; Ndzabandzaba and Hughes, 2017) and elsewhere (Zhang et al., 2018). Zhang et al. (2018) found that the aridity index was one of the most influential attributes and was well correlated with mean discharge as well as the 10th and 50th percentiles of the flow duration curve. Similarly, Ndzabandzaba and Hughes (2017) determined relationships between the aridity index and the runoff ratio, however, they also found distinct regional differences across Eswatini (Swaziland), that are less evident in the Congo River Basin. Part of this difference may be related to the substantial topographic and climate variations across the small country of Eswatini, while the much larger number of sample points (based on previous simulations) used by Ndzabandzaba and Hughes (2017), could also play a major role. Tumbo and Hughes (2015) also found regional differences in the link between aridity and hydrologic response indices, but they were not able to quantify regression relastionships, and their final result was based on simple index ranges for each identified region in the Great Ruaha River basin of Tanzania.

The main limitation for extending the predictive equations of the hydrologic indices to ungauged sub-basins is related to the spatial representativeness of the observed streamflow gauging stations. The majority of the final sub-basins used to develop the uncertainty ranges are found in region 1, even after the inclusion of the disaggregated gauging station data. While the developed uncertainty ranges of hydrologic indices can be applied with high confidence in sub-basins representing the climate and physiographic properties of region 1 (upper Lualaba and northern Oubangui), less confidence can be ascribed to their application in the other regions where some aspects of the hydrologic behavior may not have been captured. The fact that the range of the aridity index values (0.66–1.59) used to develop the predictive equations represent most of the climate variability across all 403 sub-basins of the

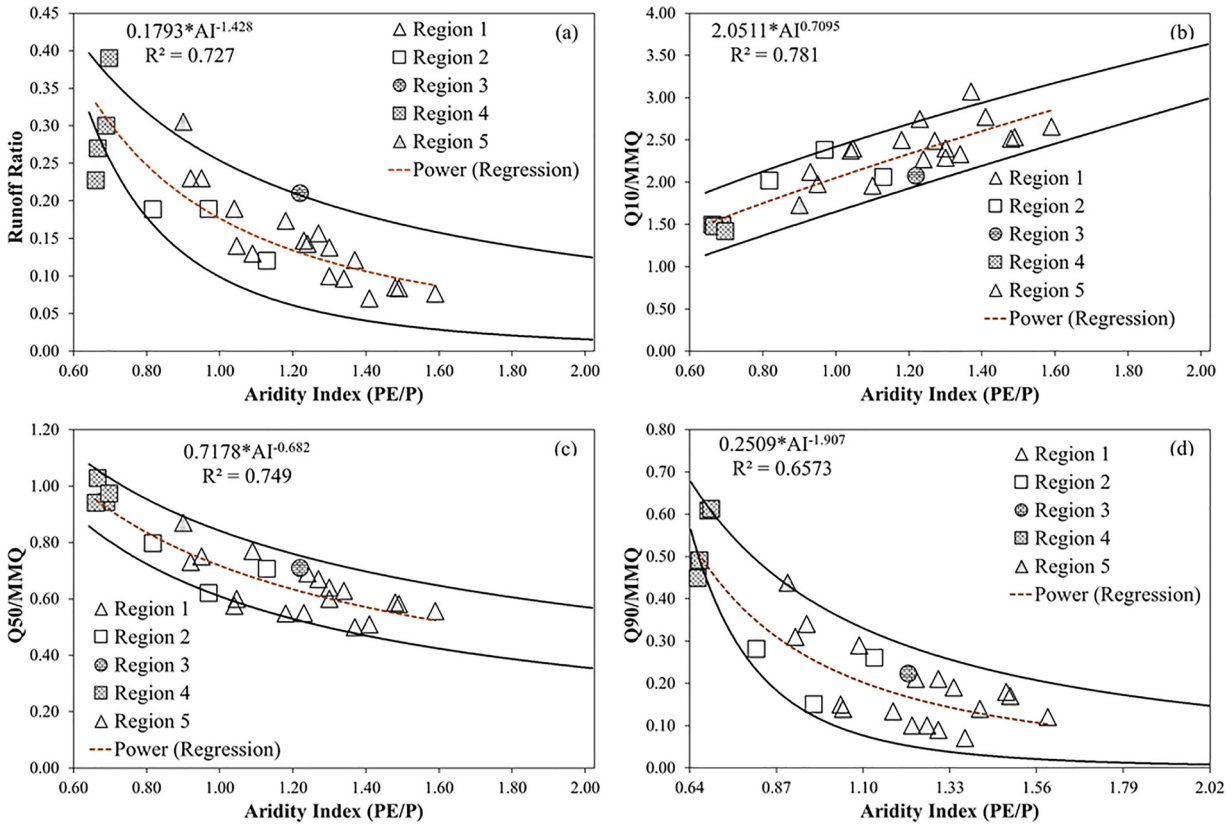

Fig. 11. Final uncertainty ranges of hydrologic indices derived based on the aridity index for all sub-basins of the Congo River Basin. (a) Runoff ratio index, (b) Q10/MMQ index, (c) Q50/MMQ index and (d) Q90/MMQ index. Regions refer to those derived from the physiographic classification of the 403 sub-basins of the Congo River Basin. The region 6 did not appear among the plotted indices because of the lack of gauged headwater sub-basins.

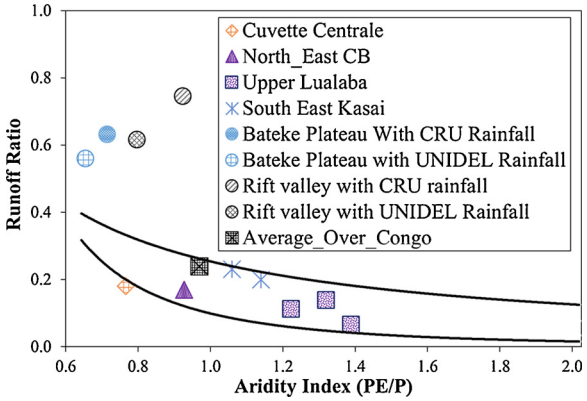

Fig. 12. Validation of the uncertainty ranges of the runoff ratio index across the Congo River Basin. The average runoff ratio observed over the entire Congo River Basin at the Kinshasa gauging station fits within the computed bounds. Gauging stations located in the Batéké plateau system (C_CB169) and rift valley system (L_CB191) are out of the computed bounds regardless of which rainfall data set is used.

Congo River Basin suggests that the approach may be quite robust in representing the diverse climate conditions within the basin. However, Fig. 12 shows some examples where some independent estimates of runoff ratio for some steep sub-basins fall well outside the uncertainty range. While these may be isolated examples of outliers, the lack of enough data seriously constrains any attempts to further validate the applicability of the relationships and the ranges of uncertainty across the whole basin.

Gnann et al. (2019) have shown that the variability of low flow in humid sub-basins of the United Kingdom and the United States could not be primarily attributed to the aridity index and that the aridity index is the key determinant of low flows only in arid regions. The Congo data tend to support this conclusion in that the uncertainty range for Q90/MMQ index is quite large and there is

no real trend in the values for Region 4 (Fig. 9d). According to Laraque et al. (1998), the hydrologic response of the sub-group (Batéké plateaux) of Region 4 sub-basins is characterized by very little seasonal variation between the low and high flows, indicating the presence of a high storage capacity groundwater system that contributes to the regulation of flows. Physical attributes describing the geological settings might have been informative in terms of including the role of sub-surface processes, but a consistent database of groundwater characteristics for the whole Congo River Basin is not yet available.

Inevitably, the developed uncertainty ranges of the hydrologic indices account for several different sources of uncertainty. These uncertainties could be due to the uncertainty in rainfall (Maidment et al., 2015; Sun et al., 2018) and evapotranspiration estimates, the length of the streamflow records, the number of the gauging stations used and their spatial distribution, the percentage of missing data, the reliability of rating curves (Kiang et al., 2018) used to convert raw stage data into streamflows, and stage observational errors. Maidment et al. (2015) found considerable differences in trend sign and magnitude (-10 and $+39$ mm yr^{-1} per decade) between different sets of global rainfall over Central Africa. They reported that the spurious negative trends identified in CRU rainfall dataset were due to the decline in rainfall gauge density across Central Africa, including the Congo River Basin. Our results have shown that substantial differences between global interpolated rainfall datasets (CRU and UNIDEL) are observed in the Batéké plateau and rift valley sub-regions located in the western and eastern parts of the Congo River Basin, respectively. It is shown that these differences were more pronounced in steep topography (Section 4.5) and constitute a common problem facing interpolated global rainfall datasets, especially in complex mountain areas, due to the limited number and spatial coverage of surface stations as well as the types of algorithms and data assimilation models used to generate the interpolated rainfall data (Sun et al., 2018). This type of uncertainty clearly affects both the aridity index and the runoff ratio.

Translating the uncertainty in instantaneous discharge observations into potential errors in monthly streamflow volumes is more difficult, particularly if the raw stage data and rating curves are not accessible, as is often the case in many countries of southern Africa. In the Congo River Basin, the lower and upper bounds of these errors for the gauging stations in the southern part of the basin (Kasai and Lualaba drainage systems) were estimated to be between -12 % and $+29$ % (Charlier, 1955 and Lempicka, 1971). The average uncertainty obtained for the Q10/MMQ and Q50/MMQ indices are less than this total value of 41 % (38 % and 32 % for Q10/MMQ and Q50/MMQ indices, respectively). These results are consistent with previous studies on the uncertainty in hydrologic signatures (Westerberg and McMillan, 2015). Westerberg et al. (2016) reported that the uncertainty in hydrologic signatures varied with signature type, with the highest uncertainties (± 30–40 %) found in high and low flow characteristics due to the uncertainty in the observed discharge and the regionalization procedures.

In the northern part of the Congo River Basin (Sangha and Oubangui drainage systems), previous studies (Laraque and Olivry, 1996; Mahé, 1995) noted a decrease in streamflow from the main tributaries of the right bank of the Congo River for the period of 1953–1993. An average 28.5 % decrease was observed in the Oubangui drainage system, 15 % in Sangha, 11 % in the Cuvette Centrale and very little change in the Batékés plateau system. Therefore, any data records that only fall within this period would be expected to generate lower mean monthly flow (MMQ) indices than would be appropriate for a longer simulation period, thus adding further levels of uncertainty. The presence of any extreme flows over the record periods, as is the case for the Congo River in the 1960s, would also impact on the MMQ indices. The majority of the gauging stations used in this study have short record periods around the 1960s. However, it is argued that the extension and infilling of missing data (Section 4.1) has at least partially overcome these effects. The uncertainty ranges of the hydrologic indices presented in this study are not affected by wetland and channel routing effects because any gauging stations located below wetlands were not represented in the 26 sub-basins used in the development of these uncertainty bounds.

The overall conclusion is that the developed relationships (and uncertainty bounds) between aridity index and the hydrologic indices are appropriate for constraining sub-basin hydrologic simulations for the whole of the Congo River Basin, with the likely exception of some areas of very steep topography on the eastern borders of the basin and the Batéké plateaux. The ultimate test of these relationships and their uncertainty bounds will be to assess the results of the constrained simulations at downstream gauging stations which have not been used in their development. It is likely that they will not work in areas downstream of the identified steep sub-basins and therefore will need to be re-calibrated for those areas.

Author statement

The Authors would like to confirm that reviewed manuscript entitled “Establishing uncertainty ranges of hydrologic indices across climate and physiographic regions of the Congo River Basin,” and associated contents have neither been published anywhere nor submitted for review to another journal. The authors would also like to confirm that the below e-mail address is accessible by the corresponding Author and is configured to accept all e-mails.

Declaration of Competing Interest

The Authors would like to confirm that there are no known conflicts of interest associated with this reviewed manuscript. The reviewed manuscript and the associated contents have neither been published anywhere nor submitted for review to another journal.

The authors acknowledge the funding support of the Congo River user Hydraulics and Morphology (CRuHM) project, which is wholly funded by The Royal Society-DFID Africa Capacity Building (RS-DFID) under the grant number “AQ150005” and also acknowledge that this financial support has no influence on the manuscript’s outcomes. I do also confirm that there are no other persons who satisfy the criteria for authorship but are not listed. I also confirm that we have given due consideration to the protection of

intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing I confirm that I have followed the regulations of my institution concerning intellectual property.

Acknowledgements

The authors acknowledge the funding support of the Congo River user Hydraulics and Morphology (CRuHM) project, which is wholly funded by The Royal Society-DFID Africa Capacity Building (RS-DFID) under the grant number “AQ150005”.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ejrh.2020.100710.

References

Addor, N., Nearing, G., Prieto, C., Newman, A.J., Le Vine, N., Clark, M.P., 2018. A Ranking of Hydrological Signatures Based on Their Predictability in Space, (i). pp. 8792–8812. https://doi.org/10.1029/2018WR022606.

Aloysius, N., Saiers, J., 2017. Simulated hydrologic response to projected changes in precipitation and temperature in the Congo River Basin. Hydrol. Earth Syst. Sci. 21 (8), 4115–4130. https://doi.org/10.5194/hess-21-4115-2017.

Alsdorf, D., Beighley, E., Laraque, A., Lee, H., Tshimanga, R., O'Loughlin, F., Mahé, G., Dinga, B., Moukandi, G., Spencer, R.G.M., 2016. Opportunities for hydrologic research in the Congo River Basin. Rev. Geophys. 54, 378–409. https://doi.org/10.1002/2016RG000517.

Batjes, N.H., 2017. Overview of Procedures and Standards in Use at ISRIC WDC-Soils. Report 2017/01. ISRIC World Soil Information, Wageningen. https://doi.org/10.17027/isric-wdcsoils.20170001.

Beck, H., De Roo, A., Van Dijk, A.I.J.M., 2015. Global maps of streamflow characteristics based on observations from several thousand catchments. J. Hydrometeorol. 16 (4), 1478–1501. https://doi.org/10.1175/jhm-d-14-0155.1.

Becker, M., Papa, F., Frappart, F., Alsdorf, D., Calmant, S., Da silva, J.S., Prigent, C., Seyler, F., 2018. Satellite-based estimates of surface water dynamics in the Congo River Basin. Int. J. Appl. Earth Obs. Geoinf. 66 (November 2017), 196–209. https://doi.org/10.1016/j.jag.2017.11.015.

Beighley, R.E., Ray, R.L., He, Y., Lee, H., Schaller, L., Andreadis, K.M., Durand, M., Alsdorf, D.E., Shum, C.K., 2011. Comparing satellite derived precipitation datasets using the Hillslope River Routing (HRR) model in the Congo River Basin. Hydrol. Process. 25 (20), 3216–3229. https://doi.org/10.1002/hyp.8045.

Bell, J.P., Tompkins, A.M., Bouka-Biona, C., Sanda, I.S., 2015. A process-based investigation into the impact of the Congo River Basin deforestation on surface climate. J. Geophys. Res. Atmos. 120, 5721–5739. https://doi.org/10.1002/2014JD022586.

Bos, A.R., Kapasa, C.K., Van Zwieten, P.A., 2006. Update on the bathymetry of Lake Mweru (Zambia), with notes on water level fluctuations. Afr. J. Aquat. Sci. 31 (1), 145–150. https://doi.org/10.2989/16085910609503882.

Buchanan, B.P., Fleming, M., Schneider, R.L., Richards, B.K., Archibald, J., Qiu, Z., Walter, M.T., 2014. Evaluating topographic wetness indices across central New York agricultural landscapes. Hydrol. Earth Syst. Sci. 18, 3279–3299. https://doi.org/10.5194/hess-18-3279-2014.

Bultot, F., 1974. Atlas climatique du bassin zaïrois. Quatrième partie: pression atmosphérique, vent en surface et en altitude, température et humidité de l'air en altitude, nébulosité et visibilité, propriétés chimiques de l'air et des précipita- tions et classifications cl. Brussels I.N.E.A.C: 193 maps.

Bwangoy, J.R.B., Hansen, M.C., Roy, D.P., De Grandi, G., Justice, C.O., 2010. Wetland mapping in the Congo River Basin using optical and radar remotely sensed data and derived topographical indices. Remote Sens. Environ. 114 (1), 73–86. https://doi.org/10.1016/j.rse.2009.08.004.

Carr, A.B., Trigg, M.A., Tshimanga, R.M., Borman, J.D., Smith, M.W., 2019. Greater Water Surface Variability Revealed by New Congo River Field Data: Implications for Satellite Altimetry Measurements of Large Rivers Geophysical Research Letters. pp. 8093–8101. https://doi.org/10.1029/2019GL083720.

Charlier, J., 1955. Études hydrographiques dans le bassin du Lualaba, Congo belge, 1952-1954, Volume 1,Partie 2 de Académie royale des sciences coloniales. Classe des sciences techniques.Mémoires in-8o. Nouv.sérVolume 1, Partie 2 de Mémoires in-80 Numéro 8 de Publications du Comité hydrographique du Bassin Congolais. pages 71. .

Clarke, K.R., 1993. Non-parametric Multivariate Analyses of Changes in Community Structure (1988). pp. 117–143.

Clausen, B., Biggs, B.J.F., 2000. Flow variables for ecological studies in temperate streams: groupings based on covariance. J. Hydrol. 237 (3–4), 184–197. https://doi.org/10.1016/S0022-1694(00)00306-1. 796.

Coynel, A., Seyler, P., Etcheber, H., Meybeck, M., Orange, D., 2005. Spatial and seasonal dynamics of total suspended sediment and organic carbon species in the Congo River. Global Biogeochem. Cycles 19 (4), 1–17. https://doi.org/10.1029/2004GB002335.

Dargie, G.C., Lewis, S.L., Lawson, I.T., Mitchard, E.T.A., Page, S.E., Bocko, Y.E., Ifo, S.A., 2017. Age, extent and carbon storage of the central Congo River Basin peatland complex. Nature 542, 86–90.

Devroey, E.J., 1951. Observations hydrographiques au Congo belge et au Ruanda-Urundi. T.VI, 3 (1951-1959). 1959. Royal Academy for Overseas Sciences. https://www.kaowarsom.be/en.

Di Prinzio, M., Castellarin, A., Toth, E., 2011. Data-driven catchment classification: application to the pub problem. Hydrol. Earth Syst. Sci. 15 (6), 1921–1935. https://doi.org/10.5194/hess-15-1921-2011.

Dyer, E.L.E., Jones, D.B.A., Nusbaumer, J., Li, H., Collins, O., Vettoretti, G., Noone, D., 2017. Congo River Basin precipitation: assessing seasonality, regional interactions, and sources of moisture. J. Geophys. Res. Atmos. 122, 6882–6898. https://doi.org/10.1002/2016JD026240.

Euser, T., Winsemius, H.C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., Savenije, H.H.G., 2013. A framework to assess the realism of model structures using hydrological signatures. Hydrol. Earth Syst. Sci. 17 (5), 1893–1912 10.5194/hess-17-1893-2013,810.

Fekete, B.M., Vo¨ro¨smarty, C.J., Grabs, W., 1999. Global, Composite Runoff Fields Based on Observed River Discharge and Simulated Water Balances, Tech. Rep. 22. Global Runoff Data Cent., Koblenz, Germany.

Gnann, S.J., Woods, R.A., Howden, N.J.K., 2019. Is There a Baseflow Budyko Curve? pp. 2838–2855. https://doi.org/10.1029/2018WR024464.

Hall, M.J., Minns, A.W., Ashrafuzzaman, A.K.M., 2002. The application of data mining techniques for the regionalisation of hydrological variables. Hydrol. Earth Syst. Sci. 6 (4), 685–694. https://doi.org/10.5194/hess-6-685-2002.

Hansen, M.C., Roy, D.P., Lindquist, E., Adusei, B., Justice, C.O., Altstatt, A., 2008. A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo River Basin. Remote Sens. Environ. 112 (5), 2495–2513. https://doi.org/10.1016/j.rse.2007.11.012.

Harris, I., Jones, P.D., Osborn, T.J., Lister, D.H., 2014. Updated high-resolution grids of monthly climatic observations – the CRU TS3. 10 Dataset 642 (May 2013), 623–642. https://doi.org/10.1002/joc.3711.

Herbst, M., Casper, M.C., 2008. Towards model evaluation and identification using Self-Organizing Maps. Hydrol. Earth Syst. Sci. 12 (2), 657–667. https://doi.org/10.5194/hess-12-657-2008.

Hughes, D.A., 2016. Hydrological modelling, process understanding and uncertainty in a southern African context: lessons from the northern hemisphere. Hydrol. Process. 30 (14), 2419–2431. https://doi.org/10.1002/hyp.10721.

Hughes, D.A., Smakhtin, V., 1996. Daily flow time series patching or extension: a spatial interpolation approach based on flow duration curves. Hydrol. Sci. J. Des Sci.

Hydrol. 41 (6), 851–871. https://doi.org/10.1080/02626669609491555.
Jothityangkoon, C., Sivapalan, M., Farmer, D.L., 2001. Process controls of water balance variability in a large semi-arid catchment: downward approach to hydrological model development. J. Hydrol. 254 (1–4), 174–198. https://doi.org/10.1016/S0022-1694(01)00496-6.
Kabantu, M.T., Tshimanga, R.M., Kileshye, O.J.M., Gumindoga, W., Beya, J.T., 2018. A GIS-based Estimation of Soil Erosion Parameters for Soil Loss Potential and Erosion Hazard in the City of Kinshasa, the Democratic Republic of Congo. pp. 51–57.
Kabuya, P.M., Hughes, D.A., Tshimanga, R.M., Trigg, M.A., Neal, J.C., 2018. Parameterization of the hydrodynamic processes of the Kamalondo wetland systems in the Upper Congo River Basin. Poster Presentation, Chapman Conference Washington DC, USA 25–27 September 2018.
Kalteh, A.M., Hjorth, P., Berndtsson, R., 2008. Review of the self- organizing map (SOM) approach in water resources: analysis, modelling and application. Environ. Model. Softw. 23, 835–845.
Kapangaziwiri, E., Hughes, D.A., Wagener, T., 2012. Incorporating uncertainty in hydrological predictions for gauged and ungauged basins in southern Africa. Africa Hydrol. Sci. J. Hydrol. Sci. J. 57 (5), 1000–1019. https://doi.org/10.1080/02626667.2012.690881.
Kiang, J.E., Gazoorian, C., Mcmillan, H., Coxon, G., Coz, J.Le., 2018. A Comparison of Methods for Stream Flow Uncertainty Estimation. https://doi.org/10.1029/2018WR022708.
Kim, D., Lee, H., Laraque, A., Tshimanga, R.M., Jung, H.C., Beighley, E., Chang, C., 2017. Mapping spatio-temporal water level variations over the central Congo River using PALSAR ScanSAR and Envisat altimetry data. Int. J. Remote Sens. 38 (23), 7021–7040. https://doi.org/10.1080/01431161.2017.1371867.
Kohonen, T., 2001. Self-Organizing Maps, third edn. Springer, Berlin, Germany.
Kult, J.M., Fry, L.M., Gronewold, A.D., Choi, W., 2014. Regionalization of hydrologic response in the Great Lakes basin: considerations of temporal scales of analysis. J. Hydrol. 519 (PB), 2224–2237. https://doi.org/10.1016/j.jhydrol.2014.09.083.
Laraque, A., Olivry, J.C., 1996. Evolution de l' hydrologie du Congo-Zaïre et de ses affluents rive droite et dynamique des transports solides et dissous. Actes de la conférence de Paris, mai 1995. IAHS Publ.no. 238, 1996.
Laraque, A., Mietton, M., Pandic, J.C.O.A., 1998. Revue des sciences de l' eau Influence des couvertures lithologiques et végétales sur les régimes et la qualité des eaux des affluents congolais du fleuve Congo. Impact of lithological and vegetation covers on flow discharge and water quality of Congolese. Revue Des Sciences de l'eau 11 (2), 209–224.
Lee, H., Beighley, R.E., Alsdorf, D., Jung, H.C., Shum, C.K., Duan, J., Guo, J., Yamazaki, D., Andreadis, K., 2011. Characterization of terrestrial water dynamics in the Congo River Basin using GRACE and satellite radar altimetry. Remote Sens. Environ. 115 (12), 3530–3538. https://doi.org/10.1016/j.rse.2011.08.015.
Lee, H., Yuan, T., Chul, H., Beighley, E., 2015. Remote Sensing of Environment Mapping wetland water depths over the central Congo River Basin using PALSAR ScanSAR, Envisat altimetry, and MODIS VCF data. Remote Sens. Environ. 159, 70–79. https://doi.org/10.1016/j.rse.2014.11.030.
Lempicka, M., 1971. Bilan hydrique du basin du fleuve Zaïre. I: Ecoulement du bassin 1950–1959. Office National de la Recherche et du Développement, Kinshasa, République Démocratique du Congo.
Ley, R., Casper, M.C., Hellebrand, H., Merz, R., 2011. Catchment classification by runoff behaviour with self-organizing maps (SOM). Hydrol. Earth Syst. Sci. 15 (9), 2947–2962. https://doi.org/10.5194/hess-15-2947-2011.
Liu, Y., Wu, X., Shen, Y., 2011. Automatic clustering using genetic algorithms. Appl. Math. Comput. 218 (4), 1267–1279. https://doi.org/10.1016/j.amc.2011.06.007.
Mahé, G., 1995. Variations des precipitations et des ecoulements en Afrique de l'Ouest et Centrale de 1951 a 1989. Secheresse- les Secheresses de Par le Monde. pp. 109–117 6 (1) (mars 1995, Bamako, Mali).
Maidment, R.I., Allan, R.P., Black, E., 2015. Recent observed and simulated changes in precipitation over Africa. Geophys. Res. Lett. 42, 8155–8164. https://doi.org/10.1002/2015GL065765.
Mayaux, P., De Grandi, G., Malingreau, J.P., 2000. Central African forest cover revisited. Remote Sens. Environ. 71 (2), 183–196. https://doi.org/10.1016/s0034-4257(99)00073-5.
McMillan, H.K., Clark, M.P., Bowden, W.B., Duncan, M., Woods, R.A., 2011. Hydrological field data from a modeller's perspective: part 1. Diagnostic tests for model structure. Hydrol. Process. 25 (4), 511–522. https://doi.org/10.1002/hyp.7841.
Mushi, C.A., Ndomba, P.M., Trigg, M.A., Tshimanga, R.M., Mtalo, F., 2019. Assessment of basin-scale soil erosion within the Congo River Basin: a review. Catena 178 (February), 64–76. https://doi.org/10.1016/j.catena.2019.02.030.
Ndehedehe, C.E., Anyah, R.O., Alsdorf, D., Agutu, N.O., Ferreira, V.G., 2019. Science of the total environment modelling the impacts of global multi-scale climatic drivers on hydro-climatic extremes (1901–2014) over the Congo River Basin. Sci. Total Environ. 651, 1569–1587. https://doi.org/10.1016/j.scitotenv.2018.09.203.
Ndzabandzaba, C., Hughes, D.A., 2017. Regional water resources assessments using an uncertain modelling approach: the example of Swaziland. J. Hydrol. Reg. Stud. 10, 47–60. https://doi.org/10.1016/j.ejrh.2017.01.002.
O'Loughlin, F., Trigg, M.A., Schumann, G.J.P., Bates, P.D., 2013. Hydraulic characterization of the middle reach of the Congo River. Water Resour. Res. 49 (8), 5059–5070. https://doi.org/10.1002/wrcr.20398.
O'Loughlin, F.E.O., Neal, J., Schumann, G.J.P., Beighley, E., Bates, P.D., 2019. A LISFLOOD-FP hydraulic model of the middle reach of the Congo. J. Hydrol.(May), 124203. https://doi.org/10.1016/j.jhydrol.2019.124203.
Olden, J.D., Poff, N.L., 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Res. Appl. 121 (September 2001), 101–121. https://doi.org/10.1002/rra.700.
Oudin, L., Kay, A., Andráassian, V., Perrin, C., 2010. Are seemingly physically similar catchments truly hydrologically similar? Water Resour. Res. 46 (11), 1–15. https://doi.org/10.1029/2009WR008887.
Peña-arancibia, J.L., Bruijnzeel, L.A., Mulligan, M., Dijk, A.I.J.M.Van., 2019. Forests as 'sponges' and 'pumps': assessing the impact of deforestation on dry-season flows across the tropics. J. Hydrol. 574 (April), 946–963. https://doi.org/10.1016/j.jhydrol.2019.04.064.
Rand, W.M., 1971. Objective criteria for the evaluation of clus- tering methods. J. Am. Stat. Assoc. 66 (336), 846–850. https://doi.org/10.2307/2284239.
Runge, J., 2007. The Congo River, Central Africa. In: Gupta, A. (Ed.), Large Rivers: Geo- Morphology and Management. John Wiley, Chichester, England, pp. 293–309.
Samba, G., Nganga, D., Mpounza, M., 2008. Rainfall and temperature variations over Congo-Brazzaville between 1950 and 1998. Theor. Appl. Climatol. 91 (1–4), 85–97. https://doi.org/10.1007/s00704-007-0298-0.
Santini, M., Caporaso, L., 2018. Evaluation of freshwater flow from rivers to the sea in CMIP5 simulations: insights from the Congo River Basin. J. Geophys. Res. Atmos. 123https://doi.org/10.1029/2017JD027422. 10,278–10,300.
Sawicz, K., Wagener, T., Sivapalan, M., Troch, P.A., Carrillo, G., 2011. Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA. Hydrol. Earth Syst. Sci. 15 (9), 2895–2911. https://doi.org/10.5194/hess-15-2895-2011.
Shafii, M., Tolson, B.A., 2015. Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. Water Resour. Res. 51, 3796–3814. https://doi.org/10.1002/2014WR016520.
Snel, M.J., 1957. Contribution à l'etude hydrogeologique du congo belge (Bull.n.7 F). Retrieved from. Academie Royale des Sciences d'outre-Mer. https://www.kaowarsom.be/fr/mémoires_en_ligne.
Sørensen, R., Zinko, U., Seibert, J., 2006. On the calculation of the topographic wetness index: evaluation of different methods based on field observations. Hydrol. Earth Syst. Sci. 10, 101–112. Retrieved from. www.copernicus.org/EGU/hess/hess/10/101/.
Somorin, O.A., Brown, C.P.H., Visseren-Hamakers, I.J., Sonwa, D.J., Arts, B., Nkem, J., 2012. The Congo River Basin forests in a changing climate: policy discourses on adaptation and mitigation (REDD+). Glob. Environ. Change 22 (1), 288–298. https://doi.org/10.1016/j.gloenvcha.2011.08.001.
Spencer, R.G.M., Stubbins, A., Hernes, P.J., Baker, A., Mopper, K., Aufdenkampe, A.K., Dyda, R.Y., Mwamba, V.L., Mangangu, A.M., Wabakanghanzi, J.N., Six, J., 2009. Photochemical Degradation of Dissolved Organic Matter and Dissolved Lignin Phenols from the Congo River 114. pp. 1–12. https://doi.org/10.1029/2009JG000968.
Spencer, R.G.M., Hernes, P.J., Ruf, R., Baker, A., Dyda, R.Y., Stubbins, A., Six, J., 2010. Temporal controls on dissolved organic matter and lignin biogeochemistry in a pristine tropical river. Democratic Republic Congo 115, 1–12. https://doi.org/10.1029/2009JG001180.
Spencer, R.G.M., Hernes, P.J., Dinga, B., Wabakanghanzi, J.N., Drake, T.W., Six, J., 2016. Origins, seasonality, and fluxes of organic matter in the Congo River. Glob.

Biogeochem. Cycles 30, 1105–1121. https://doi.org/10.1002/2016GB005427.
Srinivas, V.V., Tripathi, S., Rao, A.R., Govindaraju, R.S., 2008. Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering. J. Hydrol. 348 (1–2), 148–166. https://doi.org/10.1016/j.jhydrol.2007.09.046.
Ssegane, H., Tollner, E.W., Mohamoud, Y.M., Rasmussen, T.C., Dowd, J.F., 2012. Advances in variable selection methods II: effect of variable selection method on classification of hydrologically similar watersheds in three Mid-Atlantic ecoregions. J. Hydrol. 438–439, 26–38. https://doi.org/10.1016/j.jhydrol.2012.01.035.
Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., Hsu, K.L., 2018. A review of global precipitation data sets: data sources, estimation, and inter- comparisons. Rev. Geophys. 56, 79–107. https://doi.org/10.1002/2017RG000574.
Toth, E., 2009. Classification of hydro-meteorological conditions and multiple artificial neural networks for streamflow forecasting. Hydrol. Earth Syst. Sci. 13 (9), 1555–1566. https://doi.org/10.5194/hess-13-1555-2009.
Toth, E., 2013. Catchment classification based on characterisation of streamflow and precipitation time series. Hydrol. Earth Syst. Sci. 17 (3), 1149–1159. https://doi.org/10.5194/hess-17-1149-2013.
Tshimanga, R.M., 2012. Hydrological Uncertainty Analysis and Scenario-based Streamflow Modelling for the Congo River Basin. PhD thesis. Rhodes Univ., South Africa. http://eprints.ru.ac.za/2937/.
Tshimanga, R.M., Hughes, D.A., 2012. Climate change and impacts on the hydrology of the Congo River Basin: the case of the northern sub-basins of the Oubangui and Sangha Rivers. Phys. Chem. Earth 50–52, 72–83. https://doi.org/10.1016/j.pce.2012.08.002.
Tshimanga, R.M., Hughes, D.A., 2014. Basin-scale performance of a semidistributed rainfall-runoff model for hydrological predictions and water resources assessment of large rivers: the Congo River. Water Resour. Res. 1174–1188. https://doi.org/10.1002/2013WR014310.
Tshimanga, R.M., Bola, G.B., Kabuya, P.M., Trigg, M.A., Neal, J.C., Hughes, D.A., 2018. Towards a framework of catchment classification for hydrological predictions and water resources management in ungauged basins of the Congo River. Poster Presentation, Chapman Conference Washington DC, USA 25–27 September 2018.
Tumbo, M., Hughes, D.A., 2015. Uncertain hydrological modelling: application of the Pitman model in the Great Ruaha River basin, Tanzania. Hydrol. Sci. J. Des Sci. Hydrol. 60 (11), 2047–2061. https://doi.org/10.1080/02626667.2015.1016948.
Warton, D.I., Wright, S.T., Wang, Y., 2012. Distance-Based Multivariate Analyses Confound Location and Dispersion Effects. pp. 89–101. https://doi.org/10.1111/j.2041-210X.2011.00127.
Westerberg, I.K., Mcmillan, H.K., 2015. Uncertainty in Hydrological Signatures. pp. 3951–3968. https://doi.org/10.5194/hess-19-3951-2015.
Westerberg, I.K., Wagener, T., Coxon, G., McMillan, H.K., Castellarin, A., Montanari, A., Freer, J., 2016. Uncertainty in hydrological signatures for gauged and ungauged catchments. Water Resour. Res. 52, 1847–1865. https://doi.org/10.1002/2015WR017635.
Williams, J.R., Kannan, N., Wang, X., Santhi, C., Arnold, J.G., 2012. Evolution of the SCS runoff curve number method and its application to continuous runoff simulation. J. Hydrol. Eng. 17 (11), 1221–1229. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000529.
Wongchuig, S., Cauduro, R., De Paiva, D., Carlo, J., Collischonn, W., 2017. Multi-decadal hydrological retrospective: case study of Amazon floods and droughts. J. Hydrol. 549, 667–684. https://doi.org/10.1016/j.jhydrol.2017.04.019.
Yadav, M., Wagener, T., Gupta, H., 2007. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. Adv. Water Resour. 30 (8), 1756–1774. https://doi.org/10.1016/j.advwatres.2007.01.005.
Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J.C., Sampson, C.C., Kanae, S., Bates, P.D., 2017. A high-accuracy map of global terrain elevations. Geophys. Res. Lett. 44 (11), 5844–5853. https://doi.org/10.1002/2017GL072874.
Yang, W., Long, D., Bai, P., 2019. Impacts of future land cover and climate changes on runoff in the mostly afforested river basin in North China. J. Hydrol. 570 (January), 201–219. https://doi.org/10.1016/j.jhydrol.2018.12.055.
Yuan, T., Lee, H., Chul, H., Aierken, A., Beighley, E., Alsdorf, D.E., Tshimange, R.M., Kim, D., 2017. Remote sensing of environment absolute water storages in the Congo River floodplains from integration of InSAR and satellite radar altimetry. Remote Sens. Environ. 201 (July), 57–72. https://doi.org/10.1016/j.rse.2017.09.003.
Zeng, Z., Tang, G., Hong, Y., Zeng, C., Yang, Y., 2017. Development of an NRCS curve number global dataset using the latest geospatial remote sensing data for worldwide hydrologic applications. Remote Sens. Lett. 8 (6), 528–536. https://doi.org/10.1080/2150704X.2017.1297544.
Zhang, Z., Wagener, T., Reed, P., Bhushan, R., 2008. Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization. Water Resour. Res. 44 (12), 1–13. https://doi.org/10.1029/2008WR006833.
Zhang, Y., Chiew, Francis H.S., Li, Ming, Post, D., 2018. Predicting runoff signatures using regression and hydrological modeling approaches. Water Resour. Res. Res. 7859–7878. https://doi.org/10.1029/2018WR023325.